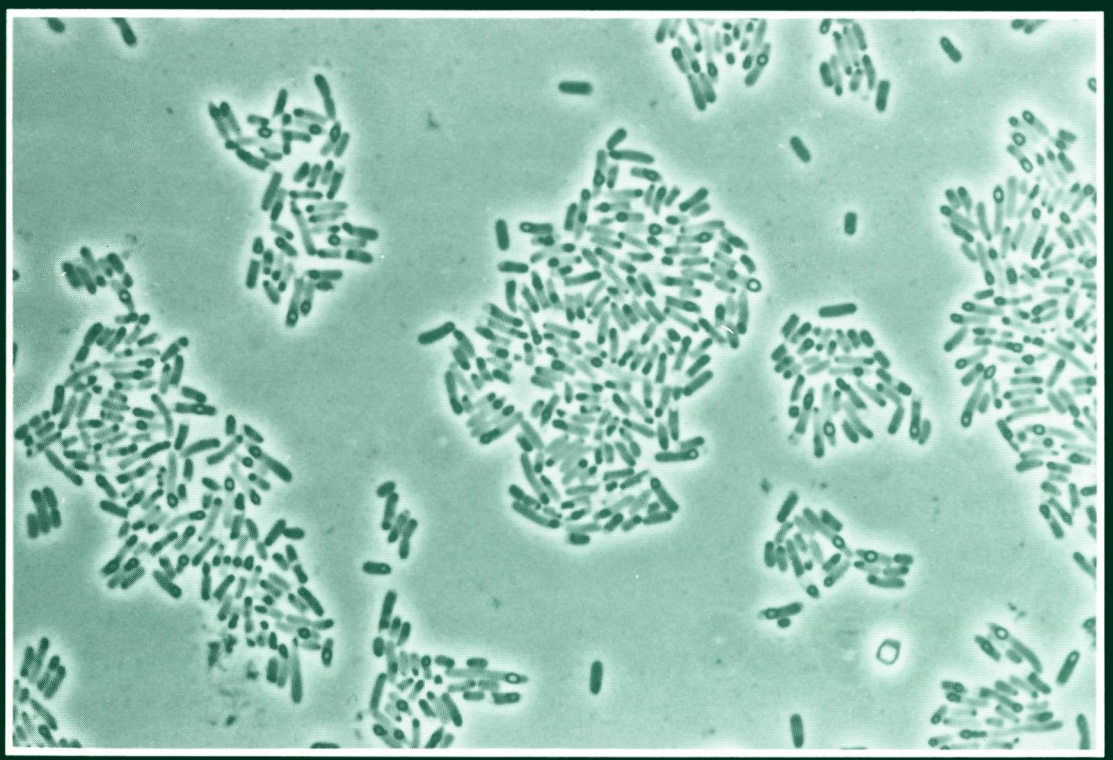
STUDIES OF THE RELATIONSHIP BETWEEN STRUCTURE AND FUNCTION OF THE SINGLE-STRANDED DNA BINDING PROTEIN OF THE FILAMENTOUS BACTERIOPHAGE M13



Fons Stassen

STUDIES OF THE RELATIONSHIP BETWEEN STRUCTURE AND FUNCTION OF THE SINGLE-STRANDED DNA BINDING PROTEIN OF THE FILAMENTOUS BACTERIOPHAGE M13

STUDIES OF THE RELATIONSHIP BETWEEN STRUCTURE AND FUNCTION OF THE SINGLE-STRANDED DNA BINDING PROTEIN OF THE FILAMENTOUS BACTERIOPHAGE M13

Een wetenschappelijke proeve op het gebied van de
Natuurwetenschappen,

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op
maandag 9 mei 1994,
des namiddags te 3.30 uur precies

door

Alphonsius Petrus Marie Stassen

geboren op 29 oktober 1958
te Heerlen.

Promotores: Prof. Dr. R.N.H. Konings
Prof. Dr. J.G.G. Schoenmakers

Co-promotor: Dr. B.J.M. Harmsen

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Stassen, Alphonsius Petrus Marie

Studies of the relationship between structure and function of the single-stranded DNA binding protein of the filamentous bacteriophage M13 / Alphonsius Petrus Marie Stassen. - [S.l. : s.n.]. - Ill.
Thesis Nijmegen. - With ref. - With summary in Dutch.
ISBN 90-9007060-5
Subjest headings: proteins / DNA / bacteriophage M13.

Cover: Phase contrast image of *Escherichia coli* cells overexpressing the Arg82-Ser gene V mutant protein. The mutant protein is accumulated in highly refractile particles, so-called 'inclusion bodies' (*Cf.* Chapter 2).

VOOR INGEBORG

VOOR MIJN MOEDER, TER NAGEDACHTENIS AAN MIJN VADER

DANKWOORD

Het werk dat tijdens een promotie-onderzoek verricht moet worden, evenals het schrijven van het proefschrift, is niet het werk van één persoon alleen. Integendeel, velen dragen er aan bij, ieder op zijn (m/v) eigen wijze. Voor wat betreft het proefschrift dat nu voor u ligt, is het niet anders en ik wil iedereen die een bijdrage aan de totstandkoming ervan heeft geleverd, van harte bedanken. Ofschoon ik hier niet iedereen kan vermelden, wil ik toch een paar namen noemen. Voor hun voortreffelijke hulp ben ik met name Jan van Deursen, Paul Folkers, Eric Schoenmakers en Guido Zaman mijn hartelijke dank verschuldigd. Daarnaast wil ik Badie Botros, John van Duynhoven, Graham Elliott, Simone Hofsteede, Edwin Janssen, Bénédicte Jenck-Soldner, Aafke Jonker, Anita Kaan, Lettie Lubsen, Hans Mesman, Carla Prinse, Clemens Prinsen, Paul Schoenmakers, Dave Schram, Gino van Teeffelen, Coen van der Veer, Els Verhoeven, Addie Weenk, Ruud Wieleman, Petra de Wilde, Cynthia Willems, Hans de Witte en Maoxiao Yu bedanken. Verder wil ik de medewerkers van het Laboratorium voor Moleculaire Biologie en het Laboratorium voor Biofysische Chemie bedanken voor de plezierige samenwerking.

En tenslotte, maar zeker niet op de laatste plaats, wil ik **INGEBORG** bedanken voor de onuitputtelijke steun die ze mij gegeven heeft en haar onvoorwaardelijk vertrouwen in mij.

FONS

CONTENTS

Chapter 1	General Introduction	9
Chapter 2	Selection and characterization of randomly produced mutants of gene V protein of bacteriophage M13	21
Chapter 3	Fluorescence studies of the binding of bacteriophage M13 gene V mutant proteins to polynucleotides	35
Chapter 4	Identification of the aromatic amino acid residues of gene V protein of bacteriophage M13, involved in single-stranded DNA binding Characterization of wild-type and mutant M13 gene V proteins by means of ^{1}H-NMR	45
Chapter 5	Nucleotide sequence of the genome of the filamentous bacteriophage I2-2 Module evolution of the filamentous phage genome	57
Chapter 6	The DNA binding wing of the single-stranded DNA binding proteins of the filamentous bacteriophages M13 and IKe: a recurrent motif in other single-stranded DNA binding proteins?	69
References		79
Summary		91
Samenvatting		93
List of publications		95
Curriculum vitae		97

CHAPTER 1

GENERAL INTRODUCTION

PROTEIN-DNA INTERACTIONS

Interactions between proteins and nucleic acids play key roles in many processes in the cell including translation, DNA packaging, DNA replication, recombination, and repair, transcription, DNA degradation and modification (Kornberg and Baker, 1992). The DNA binding proteins may interact with single-stranded or double-stranded DNA. In this chapter we shall consider mainly ssDNA binding proteins and in particular the ssDNA binding protein encoded by the filamentous bacteriophage M13.

Protein-dsDNA interactions

So far the most detailed information on DNA-protein interactions is available for proteins that bind to dsDNA in a sequence-specific manner, for instance prokaryotic and eukaryotic transcription factors or proteins involved in site-specific recombination (for reviews, see Churchill and Travers, 1991; Freemont *et al.*, 1991; Harrison, 1991; Schwabe and Rhodes, 1991). Structural data, obtained by both X-ray and NMR studies, have revealed several ubiquitous structural DNA binding motifs present in these proteins:

Helix-turn-helix : This motif consists of two α-helical structures connected by a tight β-turn resulting in an interhelical angle of about 120°. One of the two α-helices, termed the recognition helix, lies in the DNA• protein complex in the major groove. It provides sequence-specific DNA interactions.

Zinc finger: Up to now there are three types of Zinc fingers recognized, all containing Zinc as a structural element. Proteins containing a Zinc finger interact with DNA by an α-helix, the so-called recognition helix. The first type of Zinc finger, denoted C_2-H_2, consists of an antiparallel two-stranded β-sheet and an α-helix held together by one Zinc ion liganded tetrahedrally by two cysteines and two histidines. The second type of Zinc finger, designated C_2-C_2, is found in steroid and thyroid hormone receptors. It contains two loop-helix elements and two Zinc ions, but lacks a β-sheet structure. In each loop-helix element a Zinc ion is liganded to two cysteine residues situated at the beginning of the loop and two cysteine residues at the N-terminus of the α-helix. Examples of proteins containing the third type of Zinc finger are, up to now, only found in yeast transcriptional activators. The preliminary data, based on NMR and X-ray studies, show that there are two Zinc ions and six cysteine residues per DNA-binding element. Because each Zinc ion is co-ordinated by four cysteine residues, two of the six cysteines are shared.

Leucine zipper: This structural element consists of two α-helices both having a positively charged basic region at the N-terminal end. The α-helices contain four or five leucine residues periodically arranged every seventh residue. Because of this repeat, the leucine residues will lie on the same side of the α-helix, and two of these α-helices can form a dimer by adopting a parallel coiled-coil with the leucines interdigitating (the 'zipper'). The basic N-terminal regions have been shown to be involved in DNA-binding, whereas the α-helices are merely responsible for dimerization.

β-ribbon : Among the proteins that interact with DNA through a β-ribbon recognition element, are the MetJ repressor of *Escherichia coli* and the arc repressor of *Salmonella* phage P22. The β-ribbon motif consists of a two-stranded antiparallel β-sheet. In the protein-dsDNA complex the β-ribbon lies in the major groove.

This enumeration shows that dsDNA-binding is accomplished by a variety of structural motifs. However, despite the many differences, two common features can nevertheless be distinguished. Firstly, in all the motifs positively charged amino acid residues interact with the negatively charged backbone of the DNA, thus enabling non-specific interaction of dsDNA binding proteins. Secondly, the sequence-specific binding is enabled by hydrogen bonding and hydrophobic interactions between amino acid residues and functional groups mostly in the major groove of the DNA.

Protein-ssDNA interactions

Proteins that bind to ssDNA have been far less well studied than the dsDNA binding proteins. Most of what we know is related to their non-specific interaction with ssDNA. They are ubiquitously present in a variety of prokaryotic and eukaryotic organisms and include, amongst others, those encoded by filamentous phages (Alberts *et al.*, 1972; Oey and Knippers, 1972; Maeda *et al.*, 1982; Peeters *et al.*, 1983; Carne *et al.*, 1991), the phages N4 (Lindberg *et al.*, 1989), φ29 (Martín *et al.*, 1989), and T4 (Alberts *et al.*, 1968; Alberts and Frey, 1970), *E. coli* (Molineux *et al.*, 1974; Radding, 1982), yeast (Chang *et al.*, 1979), and adenovirus type 5 (Kruijer *et al.*, 1981). For the eukaryotic ssDNA binding proteins, only their biochemical properties have been characterized, and little is known about their physico-chemical properties. Only for the ssDNA binding protein SSB encoded by *E. coli* (for a review, see Meyer and Laine, 1990), the helix-destabilizing protein g32p encoded by bacteriophage T4 (for a review, see Chase and Williams, 1986), and the ssDNA binding proteins encoded by the filamentous bacteriophages M13, IKe, and Pf1, the ssDNA binding characteristics have been studied extensively in thermodynamic terms.

The ssDNA binding protein encoded by gene V (gVp) of the filamentous phage Ff (M13, f1, and fd; see Appendix) is the main subject of this thesis. For two reasons is the protein an attractive candidate to study the relationship between structure and function of a ssDNA binding protein. Firstly, the protein has been studied thoroughly, stimulated by accessibility of the system. Thus, thorough knowledge of its biological and biophysical properties are available. Secondly, because of the availability of an efficient purification procedure, and because the protein is relatively small in size, gVp lends itself for structural analysis.

The ultimate goal of our studies on gVp is to identify the amino acid residues or domains of the protein which are responsible for the execution of the respective functions of gVp. In particular we set out to examine by means of molecular biological and biophysical studies, the relationship between structure and function of gVp. We also performed a computer analysis to investigate whether the ssDNA binding motif present in gVp is, like the motifs in dsDNA binding proteins, a recurrent binding motif in other ssDNA binding proteins.

GENE V PROTEINS

The development of the DNA-cellulose affinity chromatography technique cleared the way for identification and purification of ssDNA binding proteins (Alberts and Herrick, 1971). The ssDNA binding protein of bacteriophage M13 was one of the first proteins isolated by this procedure (Alberts *et al.*, 1972; Oey and Knippers, 1972). It is classified in the group of the 'helix destabilizing proteins', because the protein lowers the melting temperature of poly d(A-T) by 20-40 °C in the absence of Mg^{2+} (Alberts *et al.*, 1972; Holwitt and Krasna, 1982). This ability to melt is caused by the strong binding of the protein to ssDNA but not to dsDNA (Alberts *et al.*, 1972; Oey and Knippers, 1972; Pratt *et al.*,1974). Later the ssDNA binding proteins of two other filamentous bacteriophages, *i.e.* IKe and Pf1, were isolated. These proteins have been characterized in both biological and biophysical terms.

In the next paragraphs we will outline the biological function of gVp encoded by phage Ff as well as the biophysical and structural characteristics of gVp encoded by phages Ff, IKe, and Pf1 (Table 1.1). For classification, structure, genetic organization and life cycle of the filamentous bacteriophages, the reader is referred to the brief survey presented in the Appendix of this chapter.

Biological function

In infected cells a pool of replicative form molecules (see Appendix) is rapidly build up and extrusion of progeny Ff phages already starts 10 min after infection to reach a steady state 30 min after infection. Moreover, the phage will remain as a parasite in the host cell indefinitely (Marvin and Hohn, 1969). In the state of persistent productive infection, a steady production of approximately 100 new phages per cell per generation is established (Webster and Lopez, 1985). To maintain both the persistent infection and the high phage production, regulation of the expression of the different Ff genes at the transcriptional as well as at the translational level, and regulation of the viral DNA replication are a prerequisite. The necessity for regulation is also illustrated by the fact that a mature phage particle contains 2750 molecules of the major coat protein and only about 5 molecules of each of the 4 different minor coat proteins.

Regulation at the transcriptional level

At the level of transcription the expression of the Ff genes is regulated by the number and strength of the promoters, the terminators, and the relative stability of the mRNAs. *In vivo* experiments have demonstrated that the Ff genome contains nine promoters, but only three terminators, and that all genes are transcribed with the same polarity as that of the viral strand (Smits *et al.*, 1978; Cashman and Webster, 1979; Cashman *et al.*, 1980; Smits *et al.*, 1980). This implies that the DNA is transcribed into a discrete number of RNA species which are initiated at different promoter sites and terminated at one of the three termination signals. The genome can be subdivided into two transcription domains separated by the two major intergenic regions (see Appendix). Transcription of the frequently transcribed domain, encompassing genes II, X, V, VII, IX,

Table 1.1 Properties of ssDNA binding proteins encoded by filamentous bacteriophages.

Protein	Mol. mass monomer (kDa)	Number of residues	Native form	Monomer copies/cell	References
M13 gVp	9.7	87	homodimer	$1 \cdot 10^5$	**a**
IKe gVp	9.8	88	homodimer		**b**
Pf1 gVp	15.4	144	homodimer		**c**

References: **a** (Oey and Knippers, 1972; Cuypers *et al.*, 1974; Nakashima *et al.*, 1974; Rasched and Pohl, 1974; Pretorius *et al.*, 1975; Cavalieri *et al.*, 1976; Hill and Petersen, 1982; Pörschke and Rauh, 1983), **b** (Peeters *et al.*, 1983), **c** (Maeda *et al.*, 1982; Kneale, 1983; Greulich *et al.*, 1985).

and VIII, is accomplished by a cascade-like mechanism in which transcripts are initiated at a number of different sites but terminated at the same ρ-independent transcription termination signal located immediately after gene VIII (Edens *et al.*, 1978; Cashman *et al.*, 1980; Grant *et al.*, 1981). Furthermore, the largest transcripts are subject to 5' post-transcriptional processing by a not yet identified host encoded endonucleolytic cleavage activity (Kokoska *et al.*, 1990). The smallest RNA species coding for genes V, VII, IX, and VIII, are relatively long-living, *i.e.* half-lives of 2.5-10 minutes, whereas the larger ones, comprising genes II through VIII, have half-lifes of less than two minutes (Cashman *et al.*, 1980). Expression of the second transcription domain which encompasses genes III, VI, I, and IV, is accomplished via a number of RNA species whose synthesis is initiated in front of genes III and IV. The transcripts are terminated at two ρ-dependent transcription termination signals located in gene I and in the major intergenic region between genes IV, and II, respectively (Smits *et al.*, 1984).

Regulation at the translational level

Three observations clearly demonstrated the existence of a regulation mechanism at the level of translation for proteins encoded by genes II, V, VII, VIII, IX, and X. Firstly, it has been shown that in contrast to the *in vivo* situation, little gXp is synthesized in an *in vitro* system (Smits *et al.*, 1978). Secondly, the high synthesis of gVp and gVIIIp reflects the abundance of their transcript. However, all mRNA species that carry coding information for gVp and gVIIIp, also carry coding information for gVIIp and gIXp which are synthesized at a much lower level. This differential translation is caused by large differences in the strength of the respective ribosome binding sites (Blumer *et al.*, 1987; Ivey-Hoyle and Steege, 1989). Furthermore, the synthesis of gVIIp and gIXp is also translationally coupled and thus highly dependent on the translation of the upstream cistron coding for gVp (Smits, 1982; Ivey-Hoyle and Steege, 1989). Thirdly, in the absence of functional gVp, or in the presence of particular gVp mutants, gIIp is overproduced (Lin and Pratt, 1972; 1974; Meyer and Geider, 1978; 1979a; Webster and Rementer, 1980; Meyer and Geider, 1981; Dotto and Zinder, 1984). *In vivo* and *in vitro* studies have unequivocally shown that this overproduction is caused by a breakdown of the gVp regulated translation of gene II mRNA (Model *et al.*, 1982; Yen and Webster, 1982; Fulford and Model, 1984; 1988a; 1988b; Zaman *et al.*, 1990). In addition, evidence has been obtained that, apart from its effect on the translation of gene II, gVp also represses at the level of translation the expression of the phage genes I, III, V, and X (Yen and Webster, 1982; Zaman *et al.*, 1991).

Regulation of DNA replication

In the initial stage of infection the pool of replicative form (RF, see Appendix) molecules increases rapidly. Later in infection the number of RF molecules is maintained at a steady state level of about 200 copies per infected cell, and progeny viral strands are synthesized. The switch from RF synthesis to the formation of progeny single-stranded viral DNA is accomplished by gVp. It inhibits RF replication by binding to the newly synthesized viral strands thus sequestering them from the DNA replication process, in which the DNA analogue of the so-called mRNA operator may be used as a nucleation site (Michel and Zinder, 1989b). Furthermore, gVp inhibits the synthesis of gIIp protein, thus slowing down RF replication. These two activities of gVp make a negative feed-back regulation feasible: An increasing number of RF molecules will lead to more gVp. This gVp will both bind to the single-stranded DNA, preventing RF formation, and it will repress the synthesis of gIIp, leading to a decreased RF replication. Conversely, if the concentration of gVp drops below a certain threshold level, the synthesis of gIIp will increase, resulting in an increase of the number of RF molecules.

For the sake of completeness, it is noted that *in vivo* experiments have shown that two more regulatory events exist (Fulford and Model, 1988a; 1988b). Firstly, apart from the function of gIIp in (+) strand synthesis (see Appendix), this protein also promotes (-) strand synthesis later in infection (Fulford and Model, 1988b). If gIIp is over-expressed, it promotes (-) strand synthesis later in infection, even in the presence of gVp. Secondly, gXp is needed for ssDNA synthesis by inhibition of gIIp.

ssDNA binding characteristics

Three major binding parameters can be used to characterize the formation of complexes between proteins and nucleic acids. Firstly, the stoichiometric number, *n*, which is the number of nucleotides covered by one protein monomer molecule. Secondly, the co-operativity parameter, ω, which defines the factor by which the binding constant increases when a protein molecule binds adjacent to an already bound protein molecule relative to isolated binding. Thirdly, the binding strength which is expressed as the intrinsic binding constant, K_{int}. K_{int} characterizes the complex formation of one protein molecule to an isolated binding site on the lattice. The effective binding constant, K_{eff}, is the product of the intrinsic binding constant and the co-operativity parameter. Various physico-chemical methods have been used to determine the binding parameters. To these methods belong measurements of ultra-violet absorption, linear or circular dichroism, and fluorescence binding studies.

Stoichiometry

The binding of phage Ff gVp and phage IKe gVp to polynucleotides is characterized by a stoichiometry of four nucleotides per protein monomer (Table 1.2). Besides this number of four, another binding mode has been reported with a different stoichiometry. ^{1}H-NMR experiments with oligonucleotides (Alma *et al.*, 1982) as well as circular dichroism titration experiments with oligonucleotides (Gray *et al.*, 1984) and with polynucleotides under low salt conditions (Kansy *et al.*,1986), revealed that under those particular conditions three nucleotides are covered by one Ff gVp monomer at complete saturation. These results have been confirmed by fluorescence binding experiments, and

Table 1.2 Binding characteristics of ssDNA binding proteins encoded by filamentous bacteriophages.

	M13 gVp		IKe gVp		Pf1 gVp	
Stoichiometry						
polynucleotide	4 ± 0.3	**a**	4 ± 0.5	**b**	4 ± 0.2	**c**
oligonucleotide	3	**d**				
Co-operativity						
polynucleotide	500 ± 200	**e**	300	**f**	100	**g**
oligonucleotide	1-5	**h**				
$-d\log(K_{eff})/d\log[NaCl]$						
polynucleotide	4 ± 0.8	**i**	3 ± 0.4	**f**		
K_{eff} at 0.2 M (salt)	$3.0 \cdot 10^5$ M^{-1} (NaCl)	**i**	$1.9 \cdot 10^5$ M^{-1} (KCl)	**f**		

References: **a** (Alberts *et al.*, 1972; Day, 1973; Anderson *et al.*, 1975; Pretorius *et al.*, 1975; Cavalieri *et al.*, 1976; Bobst *et al.*, 1982; Alma *et al.*, 1983b; Pörschke and Rauh, 1983; Bobst *et al.*,1984; Bulsink *et al.*, 1985), **b** (Peeters *et al.*, 1983; de Jong *et al.*, 1987a), **c** (Kneale *et al.*, 1982; Kneale, 1983), **d** (Alma *et al.*, 1982; 1983a; Gray *et al.*, 1984; Kansy *et al.*, 1986), **e** (Dunker and Anderson, 1975; Alma *et al.*, 1983b; Pörschke and Rauh, 1983; Bulsink *et al.*, 1985), **f** (de Jong *et al.*, 1987a), **g** (Kneale *et al.*, 1985; Carpenter and Kneale, 1991), **h** (Alma *et al.*, 1982; Bulsink *et al.*, 1986; 1988), **i** (Alma *et al.*, 1983b; Bulsink *et al.*, 1985).

fluorescence depolarization studies (Bulsink *et al.*, 1986; 1988). At a nucleotide chain length of 25 to 30 residues, both the n=3 and the n=4 mode are possible depending on the nucleotide/protein ratio (Alma *et al.*, 1983a). Two different binding modes have also been observed for the binding of phage Pf1 gVp to poly(dT). In the native nucleoprotein complex in infected cells, 3.9-4.2 nucleotides are covered by one protein monomer (Kneale *et al.*, 1982; Kneale, 1983) but *in vitro*, the binding stoichiometry was found to be 2 or 4, independent of the DNA used (Kneale *et al.*, 1985; Carpenter and Kneale, 1991): a titration experiment in which protein was added to poly(dT) resulted in n=2, whereas the reverse titration revealed n=4. These results were interpreted in terms of two distinct modes of binding. The binding site of each monomer covers 4 nucleotides. In the n=2 mode the apparent stoichiometry of 2 results from the interaction of only one subunit of the dimer with the nucleic acid. On the other hand, in the n=4 mode both monomers bind to the nucleic acid (Carpenter and Kneale, 1991).

Co-operativity

Ff gVp binds to ssDNA with a positive co-operativity (Rasched and Pohl, 1974; Cavalieri *et al.*, 1976; Alma *et al.*, 1983b). The co-operativity parameter is estimated to be about 500 ± 200 for gVp binding to polynucleotides in the n=4 mode (Dunker and Anderson, 1975; Alma *et al.*, 1983b; Pörschke and Rauh, 1983; Bulsink *et al.*, 1985). Under low ionic strength conditions, when gVp binds to polynucleotides in the n=3 mode, the co-operativity is also high (Bulsink *et al.*, 1988). In contrast, binding to oligonucleotides is hardly co-operative, ω=1-5 (Alma *et al.*, 1982; Bulsink *et al.*, 1986; 1988). The binding of IKe gVp and Pf1 gVp to polynucleotides is also highly co-operative, *i.e.* ω=300 (IKe gVp), and ω=100 (Pf1 gVp) (Peeters *et al.*, 1983; Kneale *et al.*, 1985; de Jong *et al.*, 1987a; Carpenter and Kneale, 1991).

In the infected host cell, the high positive co-operativity results in the formation of large clusters of protein molecules on the newly synthesized viral strands, resulting in immediate saturation. Thus the ssDNA strand is protected from nuclease digestion and can be packed into a new virion (Oey and Knippers, 1972; Geider, 1978).

Binding strength

Fluorescence titration experiments have shown that gVp has different affinities for various homopolynucleotides (Alma *et al.*, 1983b; Bulsink *et al.*, 1985). Circular dichroism titration experiments with poly(dA), poly(dT), poly(dC), and the alternating DNA sequences poly[d(A-C)] and poly[d(C-T)] have demonstrated that not only the base composition of the nucleic acid polymer, but the base sequence as well, is important for the binding affinity of gVp (Sang and Gray, 1989). Thus gVp shows some base or sequence specificity in the binding to polynucleotides.

The binding of Ff gVp to ssDNA is salt dependent, *i.e.* upon addition of salt the binding affinity decreases. A model for the ionic strength dependency of the effective binding constant, is described by the theory of Record *et al.* (1976; 1978):

$$\log(K_{eff}) = -(m'\psi + k) + \log[NaCl],$$

where $m'\psi$ is the number of Na^+ released from the DNA and k is the number of Cl^- released from the protein upon complex formation. From the relationship between $\log(K_{eff})$ and log[NaCl] it was postulated that in the binding process the protein loses one anion, while

three cations are released from the polynucleotide:
- $d \log(K_{eff})/d \log[NaCl] = 4 \pm 0.8$ Alma *et al.*, 1983b; Bulsink *et al.*, 1985).

Comparison of Ff gVp and IKe gVp reveals that their effective binding constants are largely similar (Peeters *et al.*, 1983; de Jong *et al.*, 1987a) (Table 1.2). The two proteins have, however, different salt dependencies of the effective binding constants. Upon binding of IKe gVp to ssDNA only three ions are released, whereas upon binding of Ff gVp to ssDNA four ions are released. This phenomenon indicates that, due to different amino acid sequences, IKe gVp has one chloride binding site less than Ff gVp (de Jong *et al.*, 1987a). Despite the sequence homology (45%) between the two proteins, it has been shown that they are not mutual co-operative in DNA binding (de Jong *et al.*, 1987a).

Structural aspects

gVp crystal structure

The successful crystallization of Ff gVp made X-ray diffraction analysis possible (McPherson *et al.*, 1976; McPherson, 1976), and a resolution to 0.23 nm was achieved (Brayer and McPherson, 1983): the molecule is non-α-helical and consists entirely of β-structures with some irregular connecting loops. Three major anti-parallel β-loops can be distinguished. The first loop is composed of residues 15 through 32, which has been termed 'DNA binding loop'. The return strand of the polypeptide chain forms a second β-loop composed of residues 33 through 49. This has been called the 'complex loop'. These two loops are arranged in such a way that they form in part a triple-stranded β-sheet. The third β-loop, the so called 'dyad loop', is composed of residues 61 through 82. This loop is oriented more or less perpendicular to the triple-stranded β-sheet. The three major β-loops project outwards from a common hydrophobic core. The dimer is found in the crystalline state and the two monomers are closely associated about a perfect dyad axis. The molecular dimensions of the dimer, which is approximately 5.5 nm x 4.5 nm x 3.5 nm in size, show that it is a compact structure. The dimer is stabilized by predominantly hydrophobic, but also by electrostatic interactions. The amino acid residues involved in these interactions are located in the six-stranded anti-parallel β-barrel that is formed by both the two extended dyad β-loops and the amino-terminal strands of both monomers. Especially, the two phenylalanine residues at position 68, located very near and parallel to the intermolecular dyad axis, are in close contact and thus their interaction might attribute to the dimer stability.

gVp structure in solution

The secondary structure of Ff gVp in solution has been studied by ^{1}H-NMR, particularly by using a missense mutant of gene V described in this thesis in which the tyrosine residue at position 41 is substituted by a histidine residue (van Duynhoven *et al.*, 1990; Folkers *et al.*, 1991). In solution the gVp structure consists of three regular secondary structure elements which are all composed of β-sheets with some connecting loops. Comparison of this structure with the crystal structure, revealed that the global folding of Ff gVp in solution and in the crystal is comparable, but at the level of the positioning of the individual amino acid residues, significant differences were found (Folkers *et al.*, 1991). The β-loop structure (residues 13 to 31) found in solution is more regular than the one found in the crystal, and in the solution structure of the DNA binding loop, the β-sheet is shifted 4 residues with respect to that of the crystal structure (van Duynhoven *et al.*, 1990). In particular, in the crystal structure residue Tyr26 is positioned at the tip of the β-loop structure, while in the solution structure this residue forms part of the anti-parallel β-sheet. This shift is propagated into the rest of the molecule. Furthermore, the N-terminal part of the protein in solution is part of a triple-stranded β-sheet, which it is not in the crystal structure. In the crystal structure the first four residues of the N-terminus point away from the bulk of the protein. Finally, there are different pairwise arrangements of the amino acids residues in the anti-parallel β sheet, which have been designated as the 'complex loop' and 'dyad loop' in the crystal structure (Brayer and McPherson, 1983; Folkers *et al.*, (1991).

The secondary structure of IKe gVp in solution has also been studied by ^{1}H-NMR (de Jong *et al.*, 1989a; van Duynhoven *et al.*, 1992;). This structure shows a high degree of similarity with that of Ff gVp.

gVp residues involved in ssDNA binding

Upon binding of Ff gVp to ssDNA, changes have been observed in the aromatic region of the proton NMR spectrum of gVp, in the circular dichroism (Day, 1973; Anderson *et al.*, 1975; Pretorius *et al.*, 1975; Gray *et al.*, 1984), ultra-violet absorbance (Day, 1973), and fluorescence emission spectra (Pretorius *et al.*, 1975). These changes indicate the presence of stacking interactions between some aromatic side chains and the nucleotide bases. To elucidate which aromatic residues are involved in ssDNA interaction, the tyrosine, histidine, and phenylalanine spin-systems have been assigned in the Ff gVp ^{1}H-NMR spectrum (Alma *et al.*, 1981a), and subsequently, a sequence specific assignment of the spin-systems was achieved that relies on crystallographic data (King and Coleman, 1987). However, in Chapter 4 of this thesis it is shown that these sequence specific assignments proved to be partially incorrect.

Upfield proton NMR shifts upon binding of gVp to oligo(dA) molecules (Coleman *et al.*, 1976; Garssen *et al.*, 1977; Coleman and Armitage, 1978; Garssen *et al.*, 1980; Alma *et al.*, 1981a; 1982; O'Connor and Coleman, 1983) and one dimensional nuclear Overhauser effects (Alma *et al.*, 1981a; 1983a) indicate that two tyrosine residues and one phenylalanine residue are involved in the binding of gVp to oligonucleotides. Two-dimensional ^{1}H-NMR revealed that Phe73 and Tyr26 are the only aromatic residues that stack significantly with the nucleotide bases (this thesis, Chapter 4; King and Coleman, 1987). Tyr26, Leu28, and Phe73' (this residue is part of the other monomer in the dimer) form the nucleotide interaction

domain. They are surrounded by an electropositive cluster, formed by Arg16, Arg21, Lys24, and Lys46, that interacts with the phosphate backbone of the DNA (King and Coleman, 1987). Tyr41 does not interact with the DNA. Its immobilization implicates a dimer-dimer contact role (King and Coleman, 1988; van Duynhoven *et al.*, 1990; Folkers *et al.*, 1991).

Models of the gVp•ssDNA complex

A computer graphics model of the helical complex of Ff gVp and ssDNA was proposed by Brayer and McPherson (1984a; 1984b; 1985). In this model the essential DNA binding element is the gVp dimer, which provides two anti-parallel DNA binding channels, each constructed from amino acids contributed by both monomers within the pair. Each channel is 1 nm wide and 4 nm long and can accommodate a fully extended DNA chain of five nucleotides in length. Upon binding the protein dimers form a right-handed helix. Complex formation is achieved in the model by stacking and electrostatic interactions. The amino acid residues that are proposed to be stacked upon the nucleic acid bases are Tyr26, Tyr34, Tyr41, and Phe73'. It is furthermore proposed that the phosphate backbone is fixed by electrostatic interactions with Arg16, Arg21, Arg80, and Lys46 (Brayer and McPherson, 1984a).

A number of studies have demonstrated that this model is not correct in a number of details. First of all, although stacking interactions are involved in complex formation, not all the bases are stacked. Electron spin resonance experiments with spin-labelled poly(dT)• gVp complexes showed that only one out of four bases is immobilized in the complex (Kao *et al.*, 1985). Also, Tyr41 is not involved in ssDNA binding (this thesis, Chapter 4), and furthermore, the channel length of 5 nucleotides is in contradiction with the stoichiometry of 4 nucleotides per protein monomer, deduced from binding experiments. In a new model for the complex (Hutchinson *et al.*, 1990), 4 nucleotides are covered by one protein monomer. However, in that model Tyr34 is involved in ssDNA binding, while it is not (this thesis, Chapter 4). Thus the second model does not fit all experimental data either. Finally, very striking is the discrepancy between the electron microscopic and neutron scattering data and the computer model. Studies with *in vitro* reconstituted complexes revealed two major structural features. Firstly, in the complex the ssDNA is found to be situated near the centre of the structure (Torbet *et al.*, 1981; Gray *et al.*, 1982a; Gray *et al.*, 1982b), and secondly, the complexes of Ff, and IKe gVp are left-handed helices (Gray, 1989). According to the model proposed by Brayer and McPherson (1984a; 1984b; 1985), the protein forms a right-handed helix and is positioned in the centre of the complex. It should be noted in passing, that the complex of phage Pf1 gVp with DNA forms a right-handed helix (Gray *et al.*, 1982a).

OUTLINE OF THIS THESIS

Aim of the studies presented in this thesis was to investigate the structure/function relationship of M13 gVp. In particular, we focussed on the elucidation of the amino acid residues involved in the different functions, *i.e.* ssDNA binding, co-operativity of binding, dimer formation, and translational regulation, of M13 gVp.

Our studies on gVp started with the construction of a library of gene V missense mutants by means of random chemical mutagenesis of gene V cloned in an expression vector. Mutant proteins were characterized both with respect to their ability to inhibit the production of phagemid DNA transducing particles and their ability to repress the translation of a chimeric *lacZ* reporter gene whose expression is controlled by the promoter and translational initiation signals of M13 gene II. The results of these studies are presented in Chapter 2.

To analyze the involvement of aromatic and basic amino acid residues in ssDNA binding, a number of missense mutants was taken from the library and the mutant proteins were purified. Subsequently the characteristics of the binding of these mutant proteins to homopolynucleotides *in vitro* were studied. The results of these experiments are presented in Chapter 3.

In Chapter 4 studies are described in which missense mutants from the library are used for the sequence-specific assignments of all aromatic amino acid residues in the gVp ^{1}H-NMR spectrum, as well as the determination of the aromatic amino acid residues involved in binding of gVp to short stretches of ssDNA. One of these missense mutants, in which the tyrosine residue at position 41 is substituted for histidine, proved to be a major break-through in ^{1}H-NMR studies performed following the studies presented in this thesis. Because of the improved solubility of this mutant protein, it has been possible to elucidate the Ff gVp secondary- and tertiary structure in solution.

Comparison of amino acid sequences of evolutionarily related proteins, which have comparable biological functions, might provide important information about the significance of particular residues or sequences with respect to the functions of the protein. For example, comparison of the amino acid sequences of the ssDNA binding proteins encoded by the evolutionarily distantly related filamentous bacteriophages Ff and IKe, has contributed valuable insights about the determinants of the ssDNA binding properties of these proteins. Because the filamentous bacteriophage I2-2 has been suggested to be a naturally occurring host range mutant of phage IKe, comparison of the amino acid sequence of its ssDNA binding protein with those of phages Ff and IKe might be very informative. Therefore, we decided to determine the nucleotide sequence of the circular single-stranded genome of phage I2-2, and compared it with the genomes of phages IKe and Ff. This is presented in Chapter 5.

^{1}H-NMR binding studies of wild-type gVp with oligonucleotides containing a covalently attached spin-label, have been used to locate the DNA binding

domains in gVp. These studies have revealed that gVp contains a β-loop structure comprising residues Arg16-Glu30 (Ff) or residues Arg16-Glu31 (IKe), which is involved in ssDNA binding. To investigate whether this ssDNA binding β-loop structure is a recurrent motif in other ssDNA binding proteins, we have compared the amino acid sequences comprising the DNA binding loops of phages Ff and IKe with the amino acid sequences of other known prokaryotic and eukaryotic ssDNA binding proteins. The result of this analysis is presented in Chapter 6.

APPENDIX

THE F-SPECIFIC FILAMENTOUS BACTERIOPHAGES

Classification and evolutionary relationship

Based on their morphology the single-stranded DNA containing bacteriophages can be subdvided into filamentous and isometric (icosahedral) phages. One of the striking differences between these two groups of phages is that the isometric phages, of which ϕX174 and G4 are the best known representatives, lyse their host cell, whereas the filamentous phages do not. Their phage particles are extruded through the cell membrane in a continuous fashion, only slightly retarding the processes of host cell growth and division. Thus far more then twenty filamentous phages are known, of which the most prominent are listed in Table 1.3. Filamentous phages belong to the smallest bacteriophages known. Their genomes are encapsulated into a thin and long (800-2000 nm, depending on the phage) protein coat (Denhardt *et al.*, 1978). Most filamentous phages can only efficiently infect cells that carry pili, which serve as their receptors. These pili are composed of filamentous polymers of identical protein subunits (pilin) generally encoded by conjugative plasmids (Marvin and Folkhard, 1986). On the basis of their pilus specificity, the filamentous phages can be subdivided into several phylogenetic classes. The phages M13, f1, and fd are variants of basically the same phage, and are collectively referred to as Ff. They only infect cells harbouring plasmids of the F-incompatibility (IncF) group. The Inc-specifity of several filamentous phages is listed in Table 1.3. Another classification of the filamentous phages is based on their helical symmetry, deduced from X-ray diffraction patterns of oriented fibers of intact phages (Marvin *et al.*, 1974a; 1974b). Class I comprises the phages Ff, ZJ/2, If1 and, IKe. They have a five-fold screw axis. Phages that lack this axis are gathered in class II. This class comprises the phages Pf1, Xf, and Pf3 (Makowski, 1984; Day *et al.*, 1988).

The various filamentous phages have not been studied to the same extent. Best studied from a physiological and genetic point of view, are Ff and IKe (for a review, see Model and Russel, 1988). Comparison of the nucleotide sequences of their genomes has demonstrated that they have evolved from a common ancestor (Beck *et al.*, 1978; van Wezenbeek *et al.*, 1980; Beck and Zink, 1981; Hill and Petersen, 1982; Peeters *et al.*, 1985). The filamentous phage I2-2 is serologically related to phage IKeh, a host range mutant of IKe (Grant *et al.*, 1978; Bradley *et al.*, 1983). This indicates that these phages are also evolutionarily related. Less extensively studied are the evolutionarily more distantly related phages Pf1 and Pf3, of which the complete nucleotide sequences have been elucidated (Luiten *et al.*, 1985; Hill *et al.*, 1991). These phages do not show any nucleotide sequence homology, nor do they show such homology with phages Ff and IKe. Recently, the complete nucleotide sequence of the *Xanthomonas* phage Cf1c, which is probably a virulent form of phage Cf1t, has been reported (Kuo *et al.*, 1991). Comparison of its nucleotide sequence with that of phage Ff has shown no homology.

Virus structure and genetic organization

The Ff bacteriophage is a long filamentous particle composed of a circular single-stranded DNA molecule, 6407 (M13, f1) or 6408 (fd) nucleotides in length, encapsulated in a flexible tubular protein sheath approximately 890 nm long and 6-7 nm wide (Denhardt *et al.*, 1978). The sheath is made up of five different phage encoded proteins. The major part of the protein sheat consists of approximately 2750 molecules of the major coat protein (the product of gene VIII) (Day and Wiseman, 1978). One end, the so-called A-end, of the particle contains about five molecules of gene III protein (Goldsmith and Konigsberg, 1977; Woolford *et al.*,

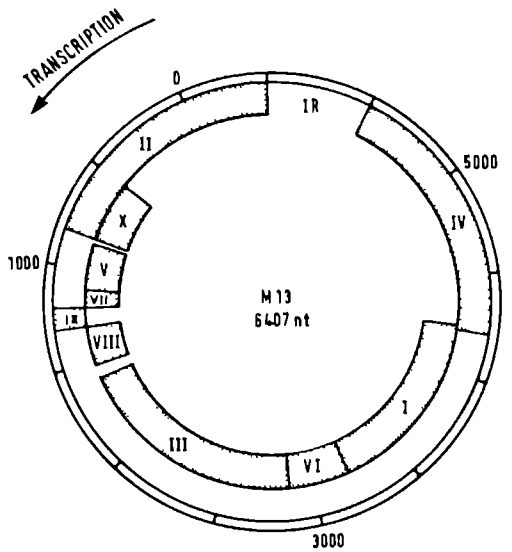


Figure 1.1 Genetic map of the genome of the filamentous bacteriophage M13.
Genes are indicated by Roman numerals. IR refers to the major intergenic region located between genes IV and II, in which both the origins for viral and complementary strand synthesis and the *cis*-acting element required for efficient phage assembly are located.

Table 1.3 Filamentous bacteriophages, their hosts and Inc-specificity.

Phage	Host	Inc-specificity	References
M13, fd, f1	*Escherichia coli*	F	**a**
ZJ/2	*Escherichia coli*	F	**b**
IKe	*Escherichia coli* *Salmonella typhimurium*	N, I2	**c**
I2-2	*Escherichia coli* *Salmonella typhimurium*	N, I2, P	**d**
If1	*Escherichia coli* *Salmonella typhimurium*	I	**e**
X	*Escherichia coli* *Salmonella typhimurium* *Serratia marcescens*	X, M, N, P-1, U, W, I2	**f**
X-2	*Escherichia coli* *Salmonella typhimurium* *Serratia marcescens*	X	**g**
tf1	*Escherichia coli* *Salmonella typhimurium* *Klebsiella oxytoca*	T	**h**
C-2	*Salmonella typhimurium* *Serratia marcescens* *Proteus mirabilis*	C	**i**
Pf1	*Pseudomonas aeruginosa*	none	**j**
Pf3	*Pseudomonas aeruginosa*	P-1	**k**
Xf	*Xanthomonas oryzae*	none	**l**
Cf	*Xanthomonas citri*	n.d.	**m**
Cf1t, Cf1c	*Xanthomonas citri*	n.d.	**n**

n.d., not determined; Inc, incompatibility

References: **a** (Loeb, 1960; Hofschneider, 1963; Marvin and Hoffmann-Berling, 1963; Zinder *et al.*, 1963), **b** (Bradley, 1964), **c** (Khatoon *et al.*, 1972), **d** (Coetzee *et al.*, 1982), **e** (Meynell and Lawn, 1968), **f** (Bradley *et al.*, 1981; Coetzee *et al.*, 1988), **g** (Coetzee *et al.*, 1988), **h** (Coetzee *et al.*, 1987), **i** (Bradley *et al.*, 1982), **j** (Takeya and Amako, 1966; Minamishima *et al.*, 1968), **k** (Stanisich, 1974), **l** (Kuo *et al.*, 1969), **m** (Dai *et al.*, 1980), **n** (Kuo *et al.*, 1987; Shieh *et al.*, 1991).

1977), and about five molecules of gene VI protein (Lin *et al.*, 1980; Grant *et al.*, 1981). The other end, the so-called B-end, contains about five molecules each of the protein molecules encoded by gene VII and gene IX (Grant *et al.*, 1981; Simons *et al.*, 1981). In the protein sheath the single-stranded phage genome is oriented in such a manner that the DNA that codes for gene III protein is located at the A-end, while that which corresponds to the major non-coding part of the genome (intergenic region, *vide infra*), at the other end (Webster *et al.*, 1981).

The genome of the Ff phage consists of ten genes which are arranged in three functional clusters (Figure 1.1) (Beck *et al.*, 1978; van Wezenbeek *et al.*, 1980; Hill and Petersen, 1982). The first cluster consists of the capsid genes (VII, IX, VIII, III, and VI). Gene III and gene VIII protein are synthesized as a precursor containing a N-terminal signal peptide (Konings *et al.*, 1975; Goldsmith and Konigsberg, 1977; Sugimoto *et al.*, 1977; Schaller *et al.*, 1978). The second cluster consists of genes I and IV which code for proteins that are indispensable for proper phage assembly (Mazur and Zinder, 1975; Wickner and Killick, 1977; Silver and Wickner, 1980). The third cluster comprises genes II, V, and X. Their protein products are of vital importance for the rolling-circle type DNA replication process of the

phage genome. The ten genes are tightly packed, *i.e.* there are very few gaps in the coding region and some genes overlap slightly at their ends. Gene X overlaps entirely with the 3' end of gene II, starting at an internal AUG which is in phase with codon 300 of gene II (Beck *et al.*, 1978; van Wezenbeek *et al.*, 1980; Hill and Petersen, 1982). Apart from the coding sequences an approximately 500 nucleotides intergenic region (IR) is present between genes II and IV (van den Hondel *et al.*, 1976; Beck *et al.*, 1978; van Wezenbeek *et al.*, 1980; Hill and Petersen, 1982). It encompasses five *cis*-acting functional elements required for proper phage DNA replication and assembly: the origins of complementary (-) and viral (+) strand synthesis (Meyer and Geider, 1979b; Geider *et al.*, 1978), the morphogenetic signal which is required for proper phage packaging (Dotto *et al.*, 1981; Dotto and Zinder, 1983), a ρ-dependent transcription terminating signal (Moses and Model, 1984), and a replication enhancer (Cleary and Ray, 1980; 1981; Dotto *et al.*, 1981; 1982). A second, but smaller intergenic region of about 50 nucleotides is located between genes III and VIII (Beck *et al.*, 1978; van Wezenbeek *et al.*, 1980; Hill and Petersen, 1982). It contains a ρ-independent transcription termination signal and a transcription promoter (Edens *et al.*, 1975; Sugimoto *et al.*, 1977).

Life cycle

A schematic representation of the life cycle of phage Ff is depicted in Figure 1.2. Three stages can be distinguished: infection, replication, and assembly.

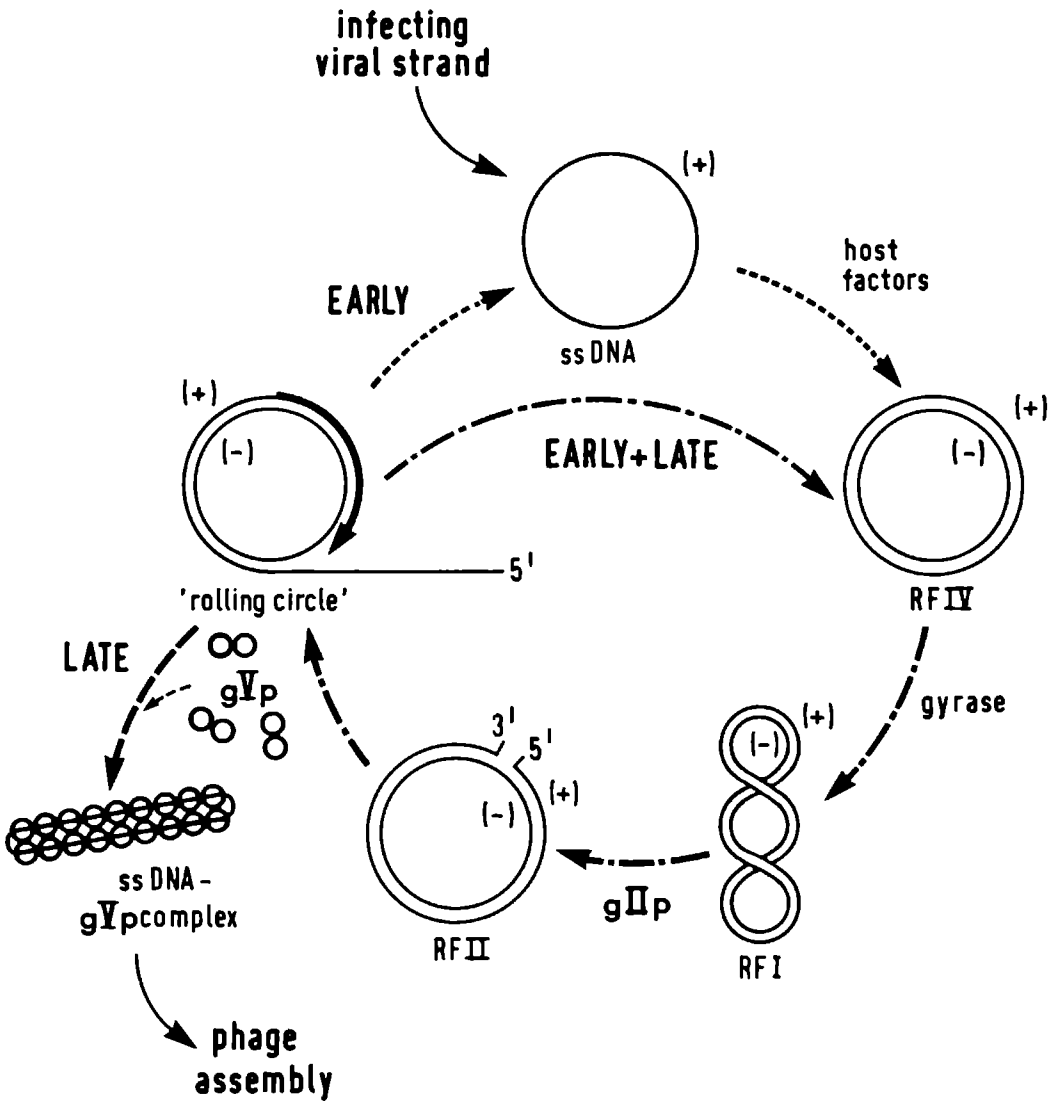


Figure 1.2 Life cycle of the filamentous bacteriophage M13.
gIIp, gene II protein; gVp, gene V protein; RF, replicative form DNA; ssDNA, single-stranded DNA; (+), viral strand; (-), complementary strand.

Infection

For absorption the Ff phage attaches to the tip of the F-pilus of the host (Caro and Schnös, 1966). Then, the phage is transported to the surface of the host cell by retraction of the pilus. At the surface, the phage genome enters the cell, while the protein molecules of the coat are deposited into the membrane (Marvin and Hohn, 1969).

Replication

Following infection, the viral DNA (SS) is replicated according to a rolling-circle type replication mechanism, in which three stages can be distinguished (Ray, 1977). The first stage consists of the conversion of the infecting viral strand via an intermediate relaxed covalently-closed double-stranded replicative form (RFIV), into a supercoiled parental double-stranded replicative form DNA (RFI): SS → RFIV → RFI. This conversion is completely dependent on host cell encoded functions, such as RNA polymerase, DNA polymerase I and III, DNA gyrase, and DNA ligase (Brutlag *et al.*, 1971; Wickner *et al.*, 1972; Geider and Kornberg, 1974; Wickner and Kornberg, 1974). Subsequent transcription and translation of the resulting RFI molecule results in the synthesis of ten phage proteins. In stage II, replication of parental RFI yields a pool of progeny RFI molecules: RFI → RFI. For this, the phage encoded gene II protein (gIIp) is absolutely required. The protein introduces a specific nick in the (+) strand origin of the viral strand of RFI, thereby creating a 3'-OH end which serves as a primer for subsequent rolling-circle type replication (Gilbert and Dressler, 1968; Fidanián and Ray, 1972; Horiuchi and Zinder, 1976). After one round of primer extension, gIIp nicks the displaced viral strand at exactly the same position and covalently closes the newly synthesized double-stranded molecule resulting in a relaxed covalently closed RF molecule, and a displaced circular single-stranded viral molecule: RFI → RFII → RFIV + SS (Horiuchi and Zinder, 1976; Horiuchi *et al.*, 1979). The RFIV molecule is converted to RFI by DNA gyrase, and can undergo the stage II replication process again, whereas early in infection the SS molecule can either enter stage I or stage III of the replication process. In stage III, which starts late after infection, viral DNA is sequestered from the replication machinery by the single-stranded DNA binding protein encoded by gene V: RFI → RFII → SS (Salstrom and Pratt, 1971; Mazur and Model, 1973; Mazur and Zinder, 1975). When the single-stranded DNA binding protein has reached a critical threshold concentration it binds specifically and co-operatively to the newly synthesized viral strand thereby preventing its conversion to RF. Thus the viral strand is made competent for assembly into progeny phage particles at the host cell membrane (Bayer, 1968; Pratt and Erdahl 1968; Salstrom and Pratt, 1971).

Assembly

In the inner membrane of the host cell, assembly of the virion occurs (Bayer, 1968; Lopez and Webster, 1985; Russel, 1991). The mechanism of this process remains obscure. It requires the activity of two phage encoded membrane proteins, *i.e.* gIp and gIVp (*vide supra*), at least one host protein, *i.e.* thioredoxin, and the morphogenetic signal (Russel, 1991). Concomitantly with the extrusion of the newly assembled virion through the membrane, the gVp molecules in the gVp•viral genome complex, are exchanged for coat protein molecules (Webster and Lopez, 1985). The released gVp molecules are re-used to sequester more newly synthesized viral strands (Pratt *et al.*, 1974).

CHAPTER 2

SELECTION AND CHARACTERIZATION OF RANDOMLY PRODUCED MUTANTS OF GENE V PROTEIN OF BACTERIOPHAGE M13

A.P.M. STASSEN, G.J.R. ZAMAN, J.M.A. VAN DEURSEN, J.G.G. SCHOENMAKERS, and R.N.H. KONINGS (1992) *Eur. J. Biochem.* **204**: 1003-1014.

ABSTRACT

Gene V protein of bacteriophage Ff (M13, f1, fd) is a master regulator of phage DNA replication and phage mRNA translation. It exerts these functions by binding to single-stranded viral DNA or to specific sequences in the 5' ends of its target mRNAs, respectively To study the structure/function relationship of gene V protein, M13 gene V was inserted in a phagemid expression vector and a library of missense and nonsense mutants was constructed by random chemical mutagenesis. Phagemids encoding gene V proteins with decreased biological activities were selected and the nucleotide sequences of their gene V fragments were determined. Furthermore, the mutant proteins were characterized both with respect to their ability to inhibit the production of phagemid DNA transducing particles and their ability to repress the translation of a chimeric *lacZ* reporter gene whose expression is controlled by the promoter and translational initiation signals of M13 gene II. From the data obtained, it can be deduced that the mechanism by which gene V protein binds to single-stranded DNA differs from the mechanism by which it binds to its target sequence in the gene II mRNA.

INTRODUCTION

The protein encoded by gene V (gVp) of the filamentous single-stranded DNA (ssDNA) bacteriophage Ff (M13, f1, fd) both regulates the replication and the expression of the viral genome (for a review, see Model and Russel, 1988). Late in infection, when its concentration has reached a certain critical threshold level, gVp switches the rolling-circle type replication of the phage genome from the production of double-stranded replicative form DNA (RF) to the asymmetric synthesis of progeny viral strands. GVp accomplishes this switch by binding in a co-operative manner to newly replicated viral strands, thereby preventing their conversion to replicative form I (RFI) and concomitantly initiating their packaging into progeny phages at the host cell membrane (Salstrom and Pratt, 1971) Besides the sequestering of viral DNA from the replication cycle, gVp also represses, among other phage genes (Zaman *et al.*, 1991), the translation of the mRNA that codes for the initiator protein of viral strand replication, *i.e.* gene II protein (gIIp), via specific binding to the 5' non-translated leader of gene II mRNA (Model *et al*, 1982; Yen and Webster, 1982, Michel and Zinder, 1989a; 1989b; Zaman *et al.*, 1990)

Since its first isolation, gVp has been an important model molecule for the study of the mechanisms by which proteins and nucleic acids interact (Alberts *et al.*, 1972; Oey and Knippers, 1972). GVp is a rather small single-stranded DNA binding protein (87 amino acids; molecular mass 9.7 kDa), that is produced in abundant amounts in the bacteriophage Ff infected cell (10^5 - 10^6 molecules per cell per generation). In solution it mainly occurs as a dimer that binds to ssDNA in a co-operative manner (Pretorius *et al.*, 1975; Cavalieri *et al*, 1976; Porschke and Rauh, 1983) The crystal structure has been examined by X-ray diffraction studies to 0.23 nm resolution (Brayer and McPherson, 1983; 1984a; 1984b) and the secondary structure of the molecule in solution has been investigated by ^{1}H-NMR techniques (van Duynhoven *et al.*, 1990; Folkers *et al.*, 1991). The protein is entirely composed of anti-parallel β-structures with some connecting loops. Comparison of the structures deduced from the ^{1}H-NMR and crystallographic data has revealed that the global folding of gVp in solution and in the crystal is similar, but that with respect to the local positions of the individual amino acids, significant differences have been found (van Duynhoven *et al.*, 1990; Folkers *et al.*, 1991).

One of the major aims of our studies on gVp is to elucidate its structure/function relationship. In particular we are interested to know the domains and amino acids

that are involved in the expression of the different functional properties of the protein. To this end, M13 gene V was placed under the control of the inducible promoter of a phagemid expression vector, and a library of gVp mutants was constructed via random chemical mutagenesis. Mutant proteins with biological properties different from that of wild-type gVp were selected by means of an *in vivo* negative genetic complementation assay. This assay is based upon the observation that the propagation of a Ff helper phage and the production of phagemid DNA transducing particles that confer resistance to ampicillin, is less efficient in cells expressing wild-type gVp than in cells expressing a biological defective gVp (*Cf.* Fulford and Model, 1988a; Terwilliger *et al.*, 1988). Clones encoding gVp molecules with a biological activity different from that of wild-type gVp were selected and the nucleotide sequences of the corresponding gene V fragments were established. Subsequently, the mutant proteins were further characterized with the aid of two functionally different *in vivo* assays. The first assay, a negative genetic complementation assay, relies on the inhibitory effect of wild-type gVp on the production of phagemid DNA transducing particles extruded from superinfected phagemid harbouring cells, under induced and non-induced conditions. In the second assay, the capability of the wild-type and mutant gVp molecules to repress the synthesis of a chimeric reporter protein, encoded by a fusion gene composed of the 5' end of M13 gene II and the 5'-truncated β-galactosidase gene of *Escherichia coli*, was measured (Zaman *et al.*, 1990).

Mutations in gVp that resulted in a decreased inhibition of the propagation of the helper phage or in a decreased translational repressor activity, were found to be randomly distributed along the gVp amino acid sequence. Comparison of the data of the two 'biological activity' assays revealed that most mutants behaved similarly in both assays. However, of some mutants the biological activity was significantly more affected in one assay than in the other. This suggests that the mechanism by which gVp binds to ssDNA differs from the mechanism by which it binds to the 5' non-translated leader of gene II mRNA.

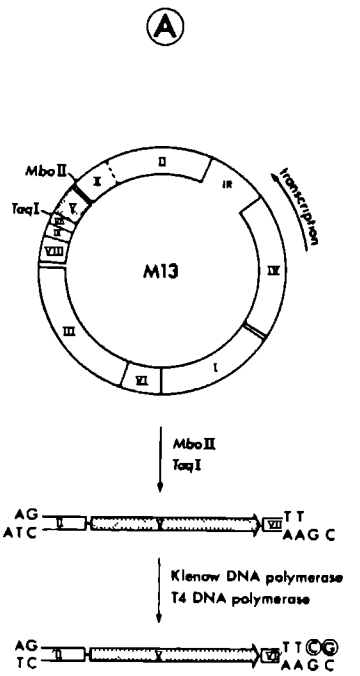


Figure 2.1 Schematic representation of the construction of the phagemids pGC2V$_{wt}$, pMS, and pMSV$_{wt}$.
(A) Procedure for the isolation of the M13 fragment containing gene V, and manipulations that were carried out to make it suitable for insertion into phagemid pKUN19. The gene V fragment comprises nucleotides 774 through 1129 of the M13 genomic map (van Wezenbeek *et al.*, 1980). The location of the phage genes is indicated by Roman numerals. IR, intergenic region. The direction of transcription is as indicated. (B) Schematic representation of the construction of phagemid pGC2V$_{wt}$ used for random chemical mutagenesis. (C) Schematic representation of the construction of the expression phagemid pMSV$_{wt}$. Only the relevant restriction sites are indicated.

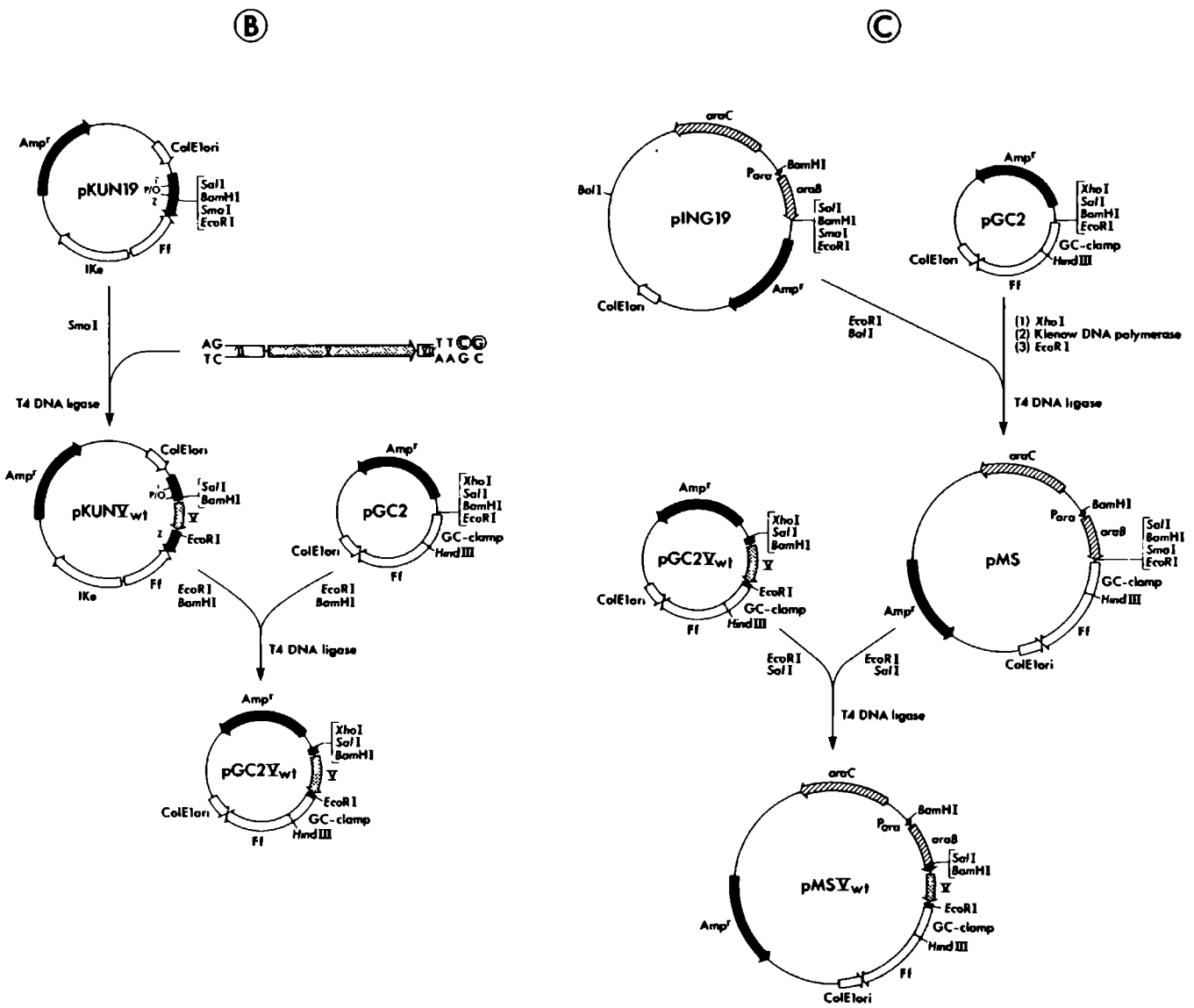
MATERIALS AND METHODS

Bacterial strains, phages, plasmids and phagemids

M13 wild-type phage and Ff helper phage R408 (Russel *et al.*, 1986) were propagated in the *Escherichia coli* strain K38 (*Hfr* (Cavalli), *tonA22*, *garB10*, *ompF627*, *relA1*, *pit-10*, *spoT1*, *metB*$^+$, *T2r*, *PO2A*, *phoA4*); recombinant plasmids and phagemids in *E. coli* JM109 (*recA1*, *endA1*, *gyrA96*, *thi*, *hsdR17*, *supE44*, *relA1*, λ$^-$,Δ(*lac-proAB*), [F', *traD36*, *proAB*, *lacI*^{q}*ZΔM15*]) or in *E. coli* MC1061[F'] (*araD139*, Δ(*ara-leu*)*7697*, Δ*lacX74*, *galU*, *galK*, *hsr*$^-$, *hsm*$^+$, *rpsL*, *rif*^{r}, [F'$_{ts}$, *Tn5*, *lac*, *kan*^{r}]). The latter strain was constructed by conjugation of *E. coli* JM83[F'] (*ara*, Δ(*lac-proAB*), *rpsL*, *φ80*, *lacZΔM15*, [F'$_{ts}$, *Tn5*, *lac*, *kan*^{r}]) and the rifampicin resistant *E. coli* strain MC1061 (*araD139*. Δ(*ara-leu*)*7697*, Δ*lacX74*, *galU*, *galK*, *hsr*$^-$, *hsm*$^+$, *rpsL*, *rif*^{r}) (Casadaban and Cohen, 1980). Single-stranded phagemid DNA transducing particles were produced via superinfection of *E. coli* MC1061[F'] cells with helper phage R408.

Plasmids and phagemids used in this study, and whose construction plus properties have been described elsewhere, are: pKUN19 (Konings *et al.*, 1987), pG2 (Johnston *et al.*, 1985), pGC1, pGC2 (Myers *et al.*, 1985), and pZII (Zaman *et al.*, 1990).

Plasmid pINg19 contains a truncated arabinose operon of *Salmonella typhimurium*, comprising the regulator gene *araC* plus its promoter and operator, and the 5' end of the *araB* gene (*araB'*) plus its promoter and operator (Figure 2.1). The *araB'* fragment is transcribed in an opposite direction as the *araC* gene. pINg19 was constructed by substitution of the gene II containing *Eco*RI-*Sal*I fragment of plasmid pG2 by the *Eco*RI-*Sal*I fragment from phagemid pKUN19 containing the multiple cloning site.



Growth media

Cells were normaly grown in 2xYT media (Miller, 1972), except for the cells used in the β-galactosidase assays. These were grown in LB medium (Miller, 1972). When appropriate, ampicillin and/or kanamycin were added at concentrations of 100 µg/ml. To induce the expression of gene V in phagemid pMSV$_{wt}$ or pMSV$_{mt}$ harbouring cells, L-arabinose was added to a final concentration of 0.4% (mass/vol.).

DNA manipulations

Basic techniques, such as plasmid/phagemid isolation, restriction endonuclease mapping, gel electrophoresis, ligation, and transformation, followed the procedures as described by Sambrook *et al.* (1989). DNA sequence analysis was performed by the chain termination method of Sanger *et al.* (1977), using the 19-base primer, 5'-dCTGGCACGCGCTGGACGCG-3', which is complementary to a GC-rich region (the GC-clamp) in the phagemids pGC1/pGC2 and pMS (Myers *et al.*, 1985).

Construction of phagemids

The procedures followed for the construction of the phagemids used for random mutagenesis and for expression studies of gene V are schematically drawn in Figure 2.1.

The mutagenesis phagemids pGC1V$_{wt}$ and pGC2V$_{wt}$ were constructed by insertion of a gene V containing *Mbo*II-*Taq*I fragment of M13 RFI (van Wezenbeek *et al.*, 1980) in the *Sma*I site of phagemid pKUN19, resulting in phagemid pKUNV$_{wt}$. In the desired orientation, gene V is flanked by a *Bam*HI site at its 5' end and an *Eco*RI site at its 3' end. Subsequently, the gene V containing *Eco*RI-*Bam*HI fragment was inserted in the phagemids pGC1 and pGC2, that had been digested with *Eco*RI and *Bam*HI. The latter two phagemids differ from each other with respect to the

orientation of their multiple cloning sites. After selection and characterization of the phagemids pGC1V_{wt} and pGC2V_{wt}, they were used for the chemical mutagenesis procedure described below. The advantage of having two recombinant phagemids with gene V in a different orientations is that upon superinfection of cells harbouring pGC1V_{wt} or pGC2V_{wt} with helper phage R408, either strand of gene V can be separately packaged into single-stranded phagemid DNA containing particles.

For the construction of the expression phagemid pMS, phagemid pGC2 and plasmid pING19 were used. The *Eco*RI-*Ball* fragment of plasmid pING19 containing the truncated operon, was inserted in phagemid pGC2, that had been digested with *Xho*I and *Eco*RI, resulting in phagemid pMS. Subsequently, the *Eco*RI-*Sal*I fragment of pGC2V_{wt} or pGC2V_{mt} (or pKUNV$_{wt}$) was inserted inbetween the *Eco*RI and the *Sal*I sites of phagemid pMS, resulting in the phagemids pMSV$_{wt}$ or pMSV$_{mt}$. In phagemid pMSV$_{wt}$, and pMSV$_{mt}$, gene V is under control of the L-arabinose inducible *araBAD* (P_{ara}) promoter (Horwitz *et al.*, 1981). In the *araB*'-gene V region, the *araB*' reading frame is continued in the 3' end of gene II. The translation of the *araB*'/'II fusion protein terminates at the same nonsense codon as the translation of gene II in phage M13. The *araB*'/'II fusion protein is composed of the first 142 amino acid residues of ribulokinase (encoded by *araB*') (Lin *et al.*, 1985), followed by 5 amino acid residues encoded by part of the pKUN19 multiple cloning site sequence, and the last 18 amino acid residues of gIIp. In phagemid pMSV$_{wt}$ the gene V containing fragment from phage M13 comprises the gene V promoter ($G_{0\ 12}$) (Smits, 1982).

Melting transition of the fragment encompassing gene V

In the chemical mutagenesis procedure used (Myers *et al.*, 1985), the conditions were chosen such that only 10% to 20% of the target fragments contain a lesion. To enrich the fragments that contain a mutation a denaturing gradient gel electrophoresis system was applied. In this system, DNA fragments migrate in an ascending denaturant concentration. At a certain denaturant concentration the DNA undergoes a helix-to-coil transition, which is accompanied by an abrupt decrease in the mobility of the DNA fragment. This is a consequence of the entanglement of the branched DNA molecule in the gel. Fragments of identical size, but differing by only one base pair, will melt at different denaturant concentrations and as a result will be locked at different positions in the gel. The resolving power of the gel is lost when the two strands of the DNA fragment are separated completely, because the single strands keep on migrating without entanglement. To prevent complete strand dissociation, a so-called GC-clamp, which is a GC-rich region that remains helical over the entire range of denaturant concentration used, was covalently coupled to the target fragment.

To determine the melting behaviour of the fragment containing wild-type gene V, 25 µg of pGC1V_{wt} RFI or pGC2V_{wt} RFI was digested with *Eco*RI and *Hin*dIII or *Bam*HI and *Hin*dIII, respectively. In both cases fragments are obtained in which gene V is covalently coupled either at its 5' (pGC1V_{wt}) or 3' end (pGC2V_{wt}) to the GC-clamp. After digestion the DNA was concentrated by ethanol precipitation and resuspended in 200 µl of E buffer (1 mM EDTA, 20 mM sodium acetate, 40 mM Tris•HCl pH 8.0) and 20 µl of loading buffer (0.25% bromophenol blue (mass/vol.), 0.25% xylene cyanol (mass/vol.), 50% glycerol (vol./vol.)). The solution was layered across the top of a 6.5% polyacrylamide gel (mass/vol.) (acrylamide/bis-acrylamide = 37.5:1), with a denaturant gradient from 0% to 80% perpendicular to the direction of electrophoresis in E buffer. 100% denaturant corresponds to 7 M urea and 40% (vol./vol.) formamide. Electrophoresis was performed at 60 °C for 5 hr and 150 V in a LKB 2001 Vertical Electrophoresis Unit. Thereafter the gel was stained with 1.5 µg/ml ethidium bromide in E buffer.

The stained patterns revealed that each fragment had only a single inflection point and consequently only one melting domain. For the gene V fragment containing the GC-clamp at its 5' end the inflection point was at 30%; while for the fragment containing the GC-clamp at its 3' end, the inflection point was at 24% denaturant. On the basis of these results, for the separation of wild-type gene V containing fragments from those containing a mutated gene V, a 10% to 40% denaturant gradient was used.

Saturation Mutagenesis

Chemical mutagenesis *in vitro* was performed essentially as described by Myers *et al.* (1985). The procedure is schematically outlined in Figure 2.2. ssDNA of phagemid pGC1V_{wt} and pGC2V_{wt} was isolated from phagemid DNA transducing particles, produced by *E. coli* MC1061[F', pGC1V_{wt}] or *E. coli* MC1061[F',pGC2V_{wt}] that had been superinfected with the helper phage R408. The ssDNA was treated with three different mutagens in such a way that, according to Myers *et al.* (1985), 10%-20% of the target fragments contain a (single) lesion. The chemical mutagens used and their potential base changes (transitions and transversions) are as follows: nitrous acid, C to T and A to G; formic acid, A and G to all bases; hydrazine, C and T to all bases.

To 40 µl of ssDNA at a concentration of 1 mg/ml, 10 µl of 2.5 M sodium acetate pH 4.3, 2 M $NaNO_2$ was added. This mixture was incubated at room temperature for 26 minutes. For the formic acid treatment 60 µl of 18 M formic acid was added to 400 µl of ssDNA at a concentration of 1 mg/ml, and incubated for 4 minutes and 20 seconds at room temperature. To mutate the DNA with hydrazine, 800 µl of 16 M hydrazine was added to 40 µl of ssDNA at a concentration of 1 mg/ml, and incubated for 4 minutes and 20 seconds at room temperature. Immediately after the incubations the DNA was concentrated by ethanol precipitation at -70 °C for 10 minutes, by the addition of 200 µl of 2.5 M sodium acetate, 100 µl of distilled H_2O, and 1 ml of 96% ethanol. The DNA was resuspended in 80 µl of TE buffer (1 mM EDTA, 10 mM Tris•HCl pH 7.6). From the chemically modified single-stranded template, double

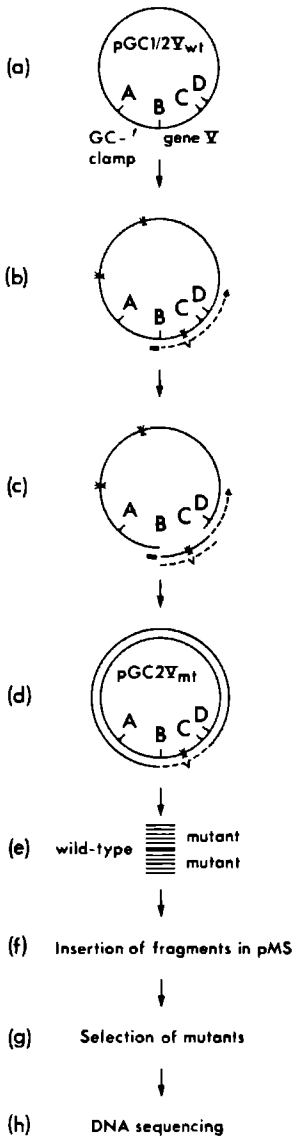


Figure 2.2 Schematic diagram of the mutagenesis procedure.
The gene V containing fragment can be excised with enzymes *B* and *C*, or *B* and *D*, and the gene V fragment plus the GC-clamp with enzymes *A* and *C*, or *A* and *D*. *A* designates a *Hin*dIII site; *B* and *C* designate *Bam*HI and *Eco*RI sites respectively for pGC1V_{wt}, and *Eco*RI and *Bam*HI sites respectively for pGC2V_{wt}; *D* designates a *Sal*I site. A single line, a double line, and a dashed line represent single-stranded DNA, double-stranded DNA, and newly synthesized DNA, respectively. The black box indicates the 19-base primer. The asterisks represent bases modified by chemical treatment.
(a) Single-stranded pGC1V_{wt} and pGC2V_{wt} are treated with mutagens. (b) The 19-base oligonucleotide primer is annealed to the single-stranded phagemids and extended with avian myeloblastosis virus reverse transcriptase. (c) The double-stranded gene V fragment is excised with enzymes *B* and *C* and inserted in non-mutagenized pGC2. (d) *E. coli* is transformed and the phagemids, which are extracted from the pool of transformants, are digested with the enzymes *A* and *D*, leaving gene V and the GC-clamp on the same DNA fragment. (e) The restriction fragments are run on a denaturing gradient gel. (f) The DNA bands above and beneath the wild-type band are excised from the gel. Then the DNA is eluted and ligated with phagemid pMS. (g) *E. coli* is tranformed and the transformants are subjected to a selection based on negative genetic complementation. (h) The nucleotide sequence of the selected mutants was established. The schematic diagram is modified from that of R.M. Myers *et al.* (1985).

stranded DNA was synthesized via annealing of the 19-base oligonucleotide primer and extending with avian myeloblastosis virus reverse transcriptase. As a consequence of the mutagenesis procedure, the primer extension products will contain misincorporations at positions corresponding to the damaged base on the single-stranded template DNA. Following extension by reverse transcriptase, the double-stranded gene V fragment was excised with *Eco*RI and *Bam*HI and inserted in non-mutagenized pGC2, that had been digested with the same restriction enzymes. After ligation and transformation of *E. coli* MC1061[F'], the transformed cells were plated on 2xYT agar plates supplemented with ampicillin. After incubation overnight, the colonies were pooled and grown overnight in 2xYT media supplemented with ampicillin. Double-stranded phagemid DNA was isolated from these cultures and digested with *Hin*dIII and *Sal*I, leaving gene V and the GC-clamp on the same DNA fragment. After phenol extraction and ethanol precipitation, the DNA was dissolved in E buffer supplemented with 10% loading buffer and subsequently loaded (approximately 5 µg of DNA per lane) on top of a 6.5% polyacrylamide gel (mass/vol.) (acrylamide/bisacrylamide = 37.5:1), with a denaturant gradient from 10% to 40% in E buffer, parallel to the direction of electrophoresis (100% denaturant defined as above). After electrophoresis, in E buffer, at 60 °C for 4.5 hour and 150 V, the gel was stained with 1.5 µl/ml ethidium bromide in E buffer. The DNA bands above and beneath the wild-type gene V band were excised from the gel and the DNA was isolated by crushing the acrylamide with a solid glass rod, and eluting overnight at 37 °C with 400 µl of elution buffer (0.5 M ammonium acetate, 0.1% SDS, 1mM EDTA). The fragments were concentrated by ethanol precipitation and inserted into phagemid pMS, which had been digested with *Hin*dIII and *Sal*I. After transformation and plating of *E. coli* MC1061[F'], appropriate gene V mutants were selected out of the pool of potential mutants by subjecting individual colonies to an *in vivo* assay based on negative genetic complementation.

Selection of gene V mutants

To identify, from a collection of random clones that contain both mutant and wild-type gene V, the mutants encoding gVp molecules with biological properties that are different from those of wild-type gVp, the single-stranded phagemid, and helper phage DNA produced upon superinfection with helper phage by induced cells harbouring the gene of interest, is compared to

that of induced cells harbouring $pMSV_{wt}$. To determine the amount of ssDNA from phagemids and helper phages extruded from infected cells, 5 µl of an overnight culture of *E. coli* MC1061[F',$pMSV_{wt}$] or *E. coli* MC1061 [F',$pMSV_{mt}$] was used to inoculate a 5 ml sample of 2xYT medium which was supplemented with of 0.4% (mass/vol.) L-arabinose. After incubation for 1.5 hr at 37 °C until early exponential phase, 10^{10} plaque forming units of the helper phage R408 (10 µl) were added and incubation was continued overnight. Cells were removed by centrifugation and the phages and transducing particles were precipitated from the supernatant by precipitation with 0.4% (mass/vol.) PEG 6000, 0.5 M NaCl. The ssDNA was isolated by phenol extraction, and analysed by electrophoresis on a 1% agarose gel.

Negative genetic complementation assay

The number of phagemid DNA transducing particles was determined by means of the negative genetic complementation assay. Two 5 ml samples of 2xYT medium, of which one was supplemented with 0.4% (mass/vol.) L-arabinose, were inoculated with 5 µl of an overnight culture of *E. coli* MC1061[F',$pMSV_{wt}$] or *E. coli* MC1061 [F',$pMSV_{mt}$], and were incubated for 1.5 hr at 37 °C until early exponential phase. Then 10^{10} plaque forming units of the helper phage R408 (10 µl) were added and incubation was continued overnight. Cells were removed by centrifugation and 100 µl of the supernatant was incubated for 3 minutes at 65 °C to kill the remaining bacteria. To measure the number of transducing particles by titration, 10 µl of appropriate dilutions of the supernatant was added to 100 µl of an overnight culture of *E. coli* MC1061[F'] and 190 µl of 2xYT medium. After incubation for 10 minutes at 37 °C (without shaking to enable the phagemids to infect the cells), the mixture was spread on 2xYT agar plates supplemented with ampicillin and incubated overnight at 37 °C. Following this incubation, the number of ampicillin resistant colonies were counted.

Translational repression assay

To monitor the translational repression activity of gVp (wild-type or mutant), β-galactosidase assays were performed as described previously (Miller, 1972; Zaman *et al.*, 1990).

Immunoreagents

M13 gene V protein was isolated as described elsewhere (Garssen *et al.*, 1977; van Duynhoven *et al.*, 1990). Polyvalent rabbit (New Zealand White) anti-gVp serum was raised by subcutaneous injection of 0.5 mg of purified gVp in 1 ml of complete Freund's adjuvant. After three weeks, the rabbit was boosted with 0.5 mg of gVp in 1 ml of incomplete Freund's adjuvant. This procedure was repeated twice at two-week intervals, and two weeks after the last booster serum was collected and the gVp antibody titer was established according to standard procedures (Harlow and Lane, 1988). Goat alkaline-phosphatase linked anti rabbit serum used to stain the immunoblots, was purchased from Zymed.

SDS-polyacrylamide gel electrophoresis and immunoblotting

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was used to determine the level of gVp synthesis encoded by M13 gene V (wild-type or mutant) carried either by phagemid pMS or by wild-type phage. To this end, two 50 ml samples of 2xYT medium, of which one was supplemented with L-arabinose, were inoculated with 100 µl of an overnight culture of *E. coli* MC1061[F'] harbouring $pMSV_{wt}$ or $pMSV_{mt}$. After incubation of the cultures at 37 °C overnight the cells were harvested and the proteins analysed by SDS-PAGE. To measure the level of production of gVp by M13 infected cells, first 50 ml of 2xYT medium, supplemented with kanamycin, was inoculated with 100 µl of an overnight culture of *E. coli* MC1061[F']. When the culture had reached a density of approximately 10^8 cells/ml, 100 µl of 1 M $CaCl_2$ and 100 µl of M13 wild-type bacteriophage (10^{12} plaque forming units/ml) were added and incubation was continued under vigorous shaking at 37 °C overnight. As a reference, uninfected cells were also grown in the same medium at 37 °C overnight. Subsequently, the cells were harvested and the proteins analysed by SDS-PAGE.

Cells, corresponding to a 2.5 ml culture with an optical density of 1.8 at 600 nm (lightpath 1 cm), were harvested by centrifugation and resuspended in 250 µl of SDS sample buffer (2% SDS (mass/vol.), 10% glycerol (vol./vol.), 5% 2-mercaptoethanol (vol./vol.), 62.5 mM Tris•HCl pH 6.8). The cells were lysed by heating for 5 minutes in a boiling waterbath. After centrifugation for 15 minutes, to remove the remainders of the cell walls, proteins were extracted from the supernatant with an equal volume of phenol saturated with TE buffer. The phenol and water phases were separated by centrifugation and the proteins were precipitated from the phenol and interphases with a 5-fold volume of acetone/0.1 M acetic acid (-20 °C for 60 minutes). The pellet was washed once with ethanol/diethylether (1:1 (vol./vol.)) and once with diethylether. Then the proteins were dissolved in 75 µl of SDS sample buffer. 10 µl of these samples were analysed on a SDS-polyacrylamide gradient (10%-20%) slab gel (Laemmli, 1970). After electrophoresis, the proteins were visualized either by staining with 0.25% (mass/vol.) Coomassie Brilliant Blue R-250 in 45% (vol./vol.) methanol, 10% (vol./vol.) acetic acid or by immunodetection. For immunodetection the proteins were transferred onto a nitrocellulose sheet by electroblotting at 5 °C for 2 hr and 90 V in a Bio-Rad Trans-blot Cell. Then the sheet was blocked for one hour at room temperature with a 3% (mass/vol.) solution of bovine serum albumin, followed by incubation for one hour with a 1/1000 dilution of rabbit polyvalent antiserum raised against gVp in a volume of 40 ml. Subsequently, the blot was incubated for one hour with 40 ml of a 1/7500 dilution of alkaline phosphatase-conjugated goat anti-rabbit serum. Finally the blot was stained by incubation at room temperature with 30 ml of a solution containing nitroblue-tetrazolium chloride (0.2 mg/ml) and 5-bromo-4-chloro-3-indolyl phosphate (0.05 mg/ml) in stainingbuffer (0.1 M NaCl, 5 mM $MgCl_2$,

0.1 M Tris•HCl pH 9.5).

RESULTS

General approach

To study the relationship between structure and function of M13 gVp, mutants of gene V have been constructed via a random chemical mutagenesis technique. To obtain all types of mutants, including those which are normally lethal to the phage, gene V was cloned in the expression phagemid pMS (Figure 2.1). Apart from the ampicillin resistance gene and the ColE1 origin of replication, this plasmid possesses several other unique properties. Firstly, gene V is placed under the control of the arabinose inducible *araBAD* promoter. Secondly, a GC-rich region lies next to gene V, which enables the separation of wild-type and mutant gene V on denaturing polyacrylamide gels, following the chemical mutagenesis procedure. Thirdly, due to the presence of the plus-strand origin and the morphogenetic signal of phage Ff, single-stranded phagemid DNA can be synthesized and packaged into phagelike particles, upon superinfection of $pMSV_{wt}$ or $pMSV_{mt}$ harbouring cells with a helper phage.

Gene V was randomly mutagenized with each of the three different mutagens nitrous acid, formic acid, and hydrazine. After incubation and complementary strand synthesis, the mutated DNA was separated from the wild-type DNA by means of electrophoresis on denaturing polyacrylamide gels (Materials and Methods). Subsequently, a negative genetic complementation assay was used to identify the gene V mutants that encode gVp with biological characteristics that are different from those of wild-type.

The biological properties of the mutants were established with the aid of two *in vivo* assays. In the first assay the number of transducing particles produced by cells that harbour gVp prior to superinfection with Ff helper phage R408, is quantitatively measured. In the second assay the translational repressor activity of wild-type and mutant gVp was determined. To verify whether, as a result of the mutagenesis procedure, the level of production of mutant gVp was still comparable to that of wild-type, the amount of gVp present in cells harbouring either phagemid $pMSV_{wt}$, or $pMSV_{mt}$ was also compared by electrophoresis on polyacrylamide gels.

Synthesis of $pMSV_{wt}$ encoded gVp

To compare the level of gVp synthesis of wild-type gene V, under control of the arabinose promoter in $pMSV_{wt}$, with the level of gVp synthesis in M13 infected cells, the proteins prepared from cell extracts were analysed by SDS-PAGE (Figure 2.3). From the banding patterns of the Coomassie Brilliant Blue stained gel and the immunoblot (Figure 2.3, lanes 3 and 5), it can be concluded that the level of gVp synthesis in cells harbouring $pMSV_{wt}$ is almost identical to that produced in phage M13 infected cells. The low level of gVp synthesis in cells containing $pMSV_{wt}$, under non-induced conditions (Figure 2.3, lane 4), is caused by constitutive synthesis from the weak gene V promoter, which also is present in this construct. The latter conclusion is supported by the observation that in the absence of L-arabinose the level of gVp synthesis, from a gene V containing fragment that is inserted in phagemid pMS in an opposite orientation as the fragment present in $pMSV_{wt}$, is identical to that instructed by $pMSV_{wt}$.

Mutagenesis and selection of gene V mutants

To obtain all theoretically possible nucleotide substitutions with the mutagens used, both strands of gene V were chemically treated. To reduce the number of multiple mutations in one target molecule, conditions were chosen such that only 10%-20% of the target molecules contain a lesion. To separate mutant fragments from the wild-type ones, electrophoresis on denaturing polyacrylamide gels was used. Eventually 534 potential mutants (including those without a gene V insert in phagemid pMS) were isolated. Gene V mutants were selected out of this pool of potential mutants by subjecting individual colonies to an *in vivo* assay based on negative genetic complementation.

The negative genetic complementation assay is based on the observation that cells containing wild-type gVp are unable to give rise to efficient phage or transducing particle production upon superinfection of these cells with a Ff helper phage (*e.g.* R408). The reason for this phenomenon is the fact that gVp interferes with the propagation of the phage. However, in cells that do not contain gVp or that produce a biologically inactive gVp, the propagation of the helper phage and the production of transducing particles is not impaired. The extent of helper phage and/or transducing particle production can both be monitored via the analysis of the amount of ssDNA extruded, and by counting of the number of transducing particles in a classical transduction experiment.

In order to screen the pool of potential mutants for real mutants, the amount of ssDNA present in the transducing particles and helper phages, produced by cells harbouring $pMSV_{mt}$ and grown in the presence of L-arabinose, was compared by agarose gel electrophoresis to the amount of ssDNA that can be isolated from transducing particles and helper phages produced by cells containing $pMSV_{wt}$. As shown in Figure 2.4 (lane 2) the amount of ssDNA that can be isolated from the culture supernatant of induced cells containing wild-type gVp, is fairly identical to the amount of ssDNA from the helper phages used for superinfection. If the induced cells contain, however, a biological inactive gVp, *e.g.* as a result of a nonsense *ochre* mutation, than the amount of single-stranded phage and transducing particle DNA that can be detected in the culture supernatant after helper phage infection is dramatically increased (Figure 2.4, lane 4). The screening of the potential mutants enabled us to subdivide the clones, apart from the clones with no insert that were not further considered, into two different classes: one consisting of clones with 'wild-type phenotype', *i.e.* resulting in very little or no ssDNA production after superinfection, and the other with '*ochre*-mutant phenotype', *i.e.* resulting in high or inter-

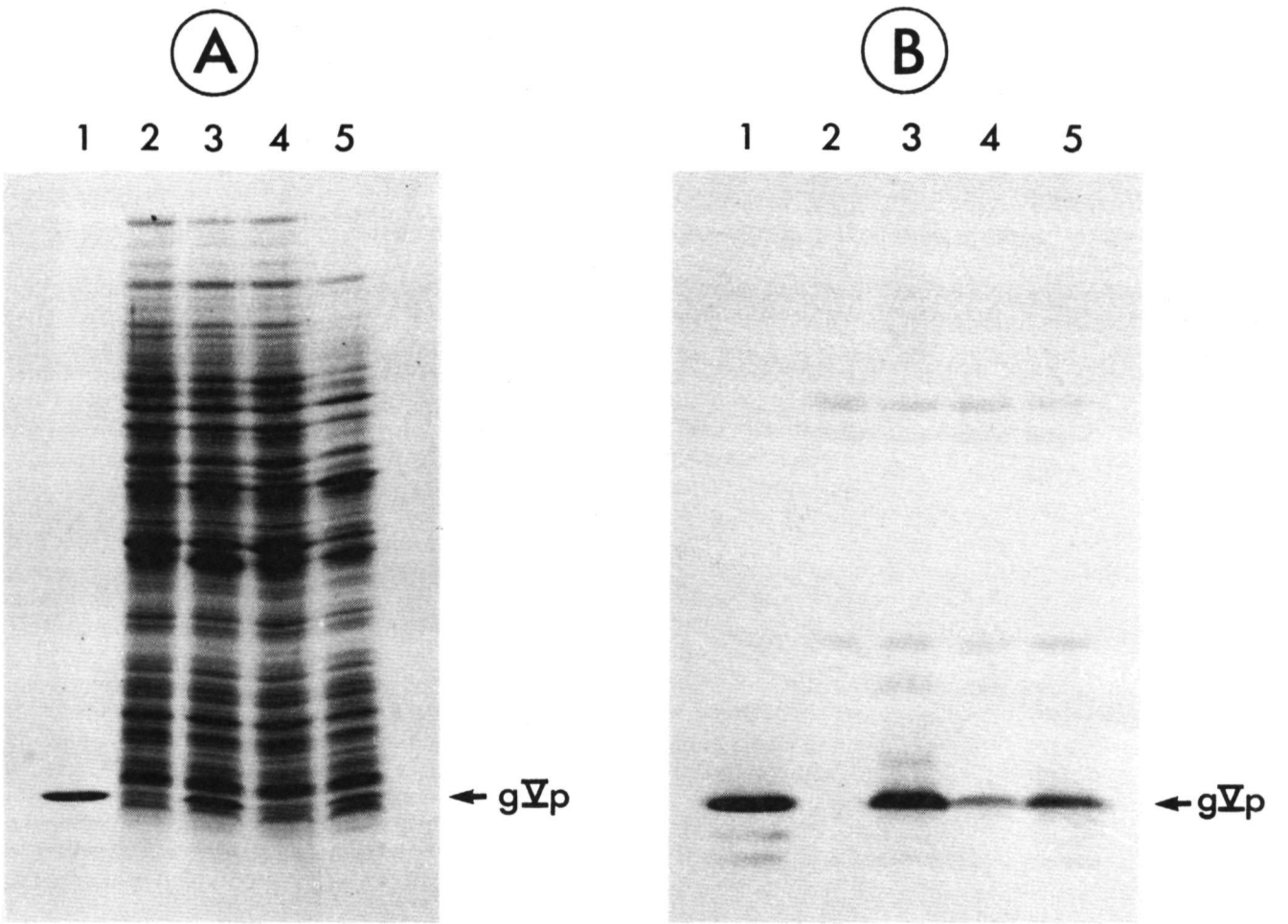


Figure 2.3 The level of wild-type gVp synthesis.
The level of gVp synthesis in wild-type bacteriophage M13 infected *E. coli* MC1061[F'] cells was compared with the level of gVp synthesis in *E. coli* MC1061[F',pMSV$_{wt}$] cells via fractionation of the proteins present in cells on SDS-PAGE. (A) Photograph of a SDS-polyacrylamide gel stained with Coomassie Brilliant Blue R-250. (B) Photograph of the immunoblot of the gel shown in A and stained with alkaline phosphatase-conjugated antibodies and 5-bromo-4-chloro-3-indolyl phosphate.
Lane 1: reference gVp. Lane 2: Proteins from *E. coli* MC1061[F']. Lane 3: Proteins from *E. coli* MC1061[F'] infected with wild-type bacteriophage M13. Lane 4: Proteins from *E. coli* MC1061[F',pMSV$_{wt}$] grown in the absence of L-arabinose. Lane 5: Proteins from *E. coli* MC1061[F',pMSV$_{wt}$] grown in the presence of 0.4% L-arabinose. An amount of protein equivalent to 0.3 ml of cell culture of OD_{600} = 1.8 (lightpath 1 cm) was loaded per lane.

mediate production of single-stranded helper phage and transducing particle DNA (Table 2.1).

To obtain mutants with silent mutations and to get an impression of the efficiency of enrichment of mutant fragment *vs.* wild-type fragment with the aid of denaturing gel electrophoresis, the nucleotide sequence of forty potential mutants with a so-called 'wild-type phenotype' was elucidated. Surprisingly, the nucleotide sequences of all of them were identical to that of wild-type gene V. From these data it thus can be concluded that the separation of wild-type and mutant fragments on the denaturing gel system used, is not absolute, for a substantial amount of the fragments in the 'mutant regions' of the gel is actually wild-type.

The potential mutants whose biological properties differ from those of wild-type gVp, according to the selection based on the ssDNA production, were further analysed with a 'packaging inhibition assay'. In this assay the amount of transducing particles produced by helper phage infected cells grown both in the absence and presence of L-arabinose is compared. As shown in Figure 2.4, cells containing phagemid pMSV$_{wt}$ are able to efficiently produce transducing particles only under non-induced conditions (lane 1), *i.e.* conditions under which fairly no wild-type gVp in the cells is present. If the cells contain, however, a prematurely terminated gVp, the amount of transducing particles produced either under induced or non-induced conditions is fairly identical (Figure 2.4, lanes 3 and 4). The number of transducing particles is determined by titration and the 'packaging inhibition activity' is defined as the ratio of the number of transducing particles extruded from cells

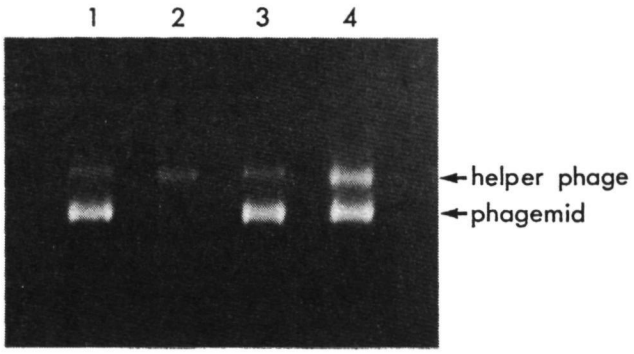


Figure 2.4 Negative genetic complementation assay.
Agarose gel analysis of the ssDNA (helper phage and phagemid) secreted by *E. coli* MC1061[F',pMSV$_{wt}$] and *E. coli* MC1061[F',pMSV$_{Gln10\text{-}ochre}$] after infection with helper phage R408.
Lane 1: Non-induced *E. coli* MC1061[F',pMSV$_{wt}$]. Lane 2: Induced *E. coli* MC1061[F',pMSV$_{wt}$]. Lane 3: Non-induced *E. coli* MC1061[F',pMSV$_{Gln10\text{-}ochre}$]. Lane 4: Induced *E. coli* MC1061[F',pMSV$_{Gln10\text{-}ochre}$].

grown in the absence of L-arabinose and the number of transducing particles extruded from the same cells grown in the presence of L-arabinose. For example *E. coli* MC1061 [F',pMSV$_{wt}$] grown in 2xYT medium supplemented with L-arabinose, produces upon superinfection about 400 times less phagemid DNA transducing particles than the same cells grown in the absence of L-arabinose.

From the mutants that had been isolated via the negative genetic complementation assay, the nucleotide sequences of the gene V fragment were established. The results are listed in Table 2.2. Forty-nine mutants with a single, and fifteen with two amino acid substitutions were identified. Nine mutants contained a nonsense codon (*amber* or *ochre*). The mutations are randomly distributed along the gVp amino acid sequence and include all types of substitutions, such as charge loss, charge gain, charge reversal, polarity reversal or without any of these changes.

Influence of amino acid substitutions in gVp on translational regulation of gene II

Apart from binding to ssDNA during regular rolling-circle type replication, gVp also represses the translation of the mRNAs of the phage genes I, II, III, V, and X (Model *et al.*, 1982; Yen and Webster, 1982; Zaman *et al.*, 1990; 1991). The strongest translational repression has been found for gene II mRNA and the nucleotide sequences involved in this translational regulation have been characterized best (Michel and Zinder, 1989a; 1989b; Zaman *et al.*, 1990).

To establish the effect of the amino acid substitutions in M13 gVp on its activity as a translational repressor, the capability of the mutant proteins to repress the translation of a fusion reporter gene consisting of the 5' end of M13 gene II, up to the 12th codon, and the 5'-truncated β-galactosidase gene of *E. coli* was measured (Zaman *et al.*, 1990). The II'/'*lacZ* reporter gene is carried on the plasmid pZII (Zaman *et al.*, 1990), which is a derivative of pACYC177 (Chang and Cohen, 1978) and is compatible with pMSV. Cells of the *lacZ* minus *E. coli* strain MC1061, containing pZII and pMSV$_{wt}$, or pMSV$_{mt}$ were grown in the absence or presence of L-arabinose, and at late exponential phase, the expression of the II'/'*lacZ* reporter gene was measured by a β-galactosidase assay (Miller, 1972; Zaman *et al.*, 1990).

The small amount of gVp present in non-induced cells harbouring pMSV$_{wt}$ (Figure 2.3) already resulted into an almost 7-fold reduction of the expression of the II'/'*lacZ* reporter gene in comparison to non-induced cells harbouring a pMSV$_{mt}$ phagemid with an *ochre* mutation at position Tyr61 (β-galactosidase activity of

Table 2.1 Phenotype of the potential mutants.
Treatment of the (+) strand (pGC2V$_{wt}$) and the (-) strand (pGC1V$_{wt}$) of gene V with each of the three mutagens, *i.e.* nitrous acid, formic acid, and hydrazine, finally resulted in the isolation of 534 potential mutants which were subdivided in clones with 'wild-type phenotype', clones with '*ochre*-mutant phenotype', and clones without an insert.

Mutagen	Strand	Number of clones			
		Total	'wild-type phenotype'	'*ochre*-mutant phenotype'	No Insert
Nitrous acid	-	157	61	73	23
	+	119	70	39	10
Formic acid	-	17	7	7	3
	+	60	29	25	6
Hydrazine	-	134	98	24	12
	+	47	34	7	6

Table 2.2 Mutants generated by the random mutagenesis method.
The library of mutants is divided into three classes: (A) Mutants with one amino acid substitution, (B) mutants with two amino acid substitutions, and (C) mutants with a nonsense codon.
GVp mutants are referred to by the wild-type residue followed by the residue number and the substituting amino acid or nonsense codon. Mutagens F, H and N designate formic acid, hydrazine and nitrous acid respectively. The 'packaging inhibition activity' and the 'translational repressor activity' are defined in the text. The level of gVp synthesis is designated physiological (P) or non-physiological (NP).

A *Mutants with a single amino acid substitution*

Mutant	Mutagen	'Packaging inhibition activity'	'Translational repressor activity'	Level of gVp synthesis
wild-type	-	407	42	P
Ile2-Val	H	20	2	P
Ile2-Thr	N	49	11	NP
Lys3-Glu	N	136	23	NP
Glu5-Lys	N	41	8	P
Ile6-Val	N	939	33	P
Ile6-Thr	N	2	10	NP
Lys7-Arg	H	15	19	P
Pro8-Thr	N	149	40	NP
Ser9-Phe	N	10	6	P
Gln10-Arg	N	4	2	NP
Gln10-Leu	F	3	6	NP
Ala11-Thr	N	6	8	NP
Gln12-Arg	N	84	33	P
Phe13-Leu	N	4	3	P
Thr14-Ala	N	74	14	P
Thr15-Ala	N	19	7	NP
Arg16-Cys	N	2	5	NP
Arg16-His	H	12	3	NP
Gly18-Asp	N	7	2	NP
Gly23-Cys	F	5	3	NP
Gly23-Asp	N	2	3	NP
Tyr26-Cys	N	32	9	NP
Cys33-Arg	N	3	3	NP
Tyr34-His	N	22	7	NP
Asp36-Asn	N	48	4	NP
Asp36-Gly	N	6	4	NP
Gly38-Asp	N	22	4	NP
Asn39-Ile	F	135	4	NP
Tyr41-His	N	6	3	P
Tyr41-Phe	F	7	7	P
Pro42-Ser	N	8	7	NP
Leu44-His	F	3	6	NP
Ile47-Thr	N	95	3	P
Leu49-Ile	H	15	6	NP
Leu49-Phe	N	2	4	NP
Ala55-Val	N	9	6	NP
Ala57-Thr	N	3	2	NP
Pro58-Ser	N	211	4	NP
Pro58-Leu	N	2	3	NP
Gly59-Ser	N	4	5	NP
Tyr61-His	N	10	5	NP
Tyr61-Ser	F	10	3	NP
Ser66-Pro	N	2	7	NP
Ser67-Phe	N	1	5	NP
Phe68-Leu	N	1	1	P
Val70-Ala	N	6	6	NP
Phe73-Leu	N	4	2	NP
Arg82-Ser	F	3	2	NP
Val84-Ala	N	6	3	NP

B *Mutants with two amino acid sustitutions*

Mutant	Mutagen	'Packaging inhibition activity'	'Translational repressor activity'	Level of gVp synthesis
Ile2-Thr Leu37-Ser	N	4	3	NP
Ile2-Val Tyr41-Cys	N	4	4	NP
Ile2-Val Leu76-Phe	N	2	1	NP
Ala11-Thr Ile47-Thr	N	3	3	NP
Ala11-Val Tyr56-Cys	N	4	2	NP
Ala11-Thr Ala57-Thr	N	3	5	NP
Ala11-Val Ala86-Val	N	3	6	NP
Gln12-Lys Tyr26-Asn	F	5	6	NP
Gly18-Ser Gly38-Ser	H	4	3	NP
Gly18-Asp Tyr56-His	N	6	3	NP
Gly23-Asp Leu37-Ser	N	2	5	NP
Pro25-Ser Tyr56-Cys	N	4	2	NP
Leu28-Pro Ala57-Thr	N	6	5	NP
Leu37-Ser Phe68-Ser	N	2	2	NP
Ala55-Val Leu76-Phe	N	4	6	NP

C *Mutants with a nonsense codon*

Mutant	Mutagen	'Packaging inhibition activity'	'Translational repressor activity'	Level of gVp synthesis
Gln10-*ochre*	N	2	2	P
Gln10-*ochre* Pro25-Ser Ala86-Val	N	2	1	P
Gln12-*ochre*	F	2	1	P
Lys24-*amber*	F	2	1	P
Ser27-*ochre*	F	3	2	P
Glu30-*amber*	F	2	1	P
Asp36-Tyr Lys46-Asn Glu51-*ochre*	F	2	1	NP
Glu51-*ochre*	F	2	1	NP
Tyr61-*ochre* Thr62-Ala	H	2	1	NP

non-induced cells harbouring pZII + $pMSV_{wt}$ is 3000 Miller units; of non-induced cells harbouring pZII + $pMSV_{Tyr61\text{-}ochre}$: 20000 Miller units; *Cf.* Zaman *et al.*, 1990). When the cells were grown in the presence of inducer, the β-galactosidase activity of cells harbouring $pMSV_{wt}$ was further decreased by a factor of 7 (to 430 Miller units), whereas that of induced cells harbouring $pMSV_{Tyr61\text{-}ochre}$ was virtually unaffected (18000 Miller units). To compensate for the repression in the absence of inducer, the 'translational repressor activity' was defined as the ratio of the β-galactosidase activity of induced cells harbouring pZII plus $pMSV_{Tyr61\text{-}ochre}$ (18000 Miller units), and induced cells harbouring pZII plus $pMSV_{wt}$, or pZII plus any of the other $pMSV_{mt}$ constructs (*e.g.* the 'translational repressor activity' of wild-type gVp is 18000/430 = 42).

The 'translational repressor activities' of the mutants are listed in Table 2.2. All mutants tested had an activity that was lower than that of wild-type gVp.

Level of gVp synthesis

Besides measuring the 'packaging inhibition activity' and the 'translational repressor activity' of wild-type and mutant gVp, their level of synthesis under non-induced and induced conditions was also studied. This was prompted by the observation that cells of which the mutant proteins (*e.g.* Tyr26-Cys and Tyr61-His) were to be purified for investigation of their binding characteristics, harboured huge amounts of gVp as revealed by SDS-PAGE. To be able to ascribe decreased biological activities to missense mutations and not to a reduced protein concentration in the cell, the level of gVp synthesis of all the mutants was established.

The level of mutant gVp synthesis was compared with the level of wild-type gVp synthesis via fractionation of the proteins present in non-induced and induced cells on SDS-PAGE followed by staining with Coomassie Brilliant Blue R-250. To our surprise, fifty-five of the seventy-three mutants showed a very high level of gVp synthesis. These fifty-five mutants include thirty-seven out of the forty-nine (76%) mutants with a single amino acid substitution, all the mutants with two amino acid substitutions, and three of the nine mutants with a premature nonsense codon. Densitometric scanning of the lanes of the SDS-PAGE gels of the mutants with the extaordinarily high gVp synthesis, revealed that at least 10% of the total cellular protein consisted of mutant gVp. This is about 10 times more than the amount of gVp produced in cells harbouring the phagemid $pMSV_{wt}$. In Figure 2.5 mutant Arg82-Ser which is representative for the mutants with a high gVp synthesis, is compared with wild-type gVp. The high level of gVp production is neither caused by mutations in the gene V promoter, nor by mutations in the Shine-Dalgarno sequence. Attempts to purify gVp from cells with a very high production level of gVp has revealed that in almost all cases the protein is accumulated in these cells in highly refractile particles, so-called 'inclusion bodies', as seen by means of phase contrast microscopy (A.P.M. Stassen and R.N.H. Konings, unpublished results). The formation of these 'inclusion bodies' is usually observed in prokaryotic cells when eukaryotic or bacterial proteins are overproduced (*i.e.* produced in much larger amounts than under natural circumstances) (Marston, 1986; Schein, 1989). It is generally believed that 'inclusion bodies' are composed of intermediates in protein folding pathways (Mitraki and King, 1989). Probably 'inclusion body' deposition is the result of incorrect or inefficient folding resulting in hydrophobic aggregation of folding intermediates.

However, SDS-PAGE analyses revealed that not all gVp molecules of the 'inclusion body' forming mutants were insoluble, but that a significant fraction was present as soluble protein in the cytoplasmatic cell fraction (data not shown). Furthermore, the observation that the 'inclusion body' forming mutants studied, *e.g.* Lys3-Glu and Pro8-Thr, are both able to efficiently repress the expression of the II'/'Z reporter gene and to inhibit the production of phagemid DNA transducing particles (Table 2.2), demonstrates that the mutant protein molecules present in the soluble fraction are biological active.

On the basis of the level of gVp production the mutant proteins can be subdivided into two catagories. The first category comprises the mutants with a production level that is comparable to that of wild-type gVp. This production is designated as physiological (P). The second category comprises the mutants with a very high production level. The production level of these 'inclusion body' forming mutants is designated as non-physiological (NP). The results are summerized in Table 2.2.

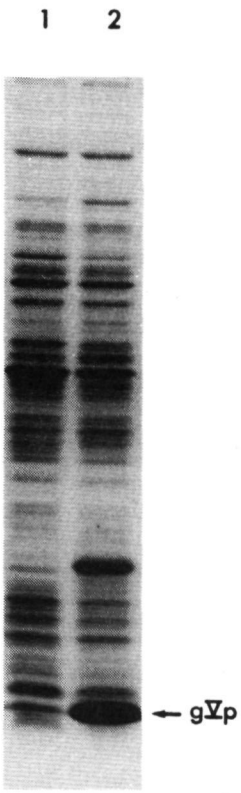


Figure 2.5 Overexpression of mutant Arg82-Ser gVp.
Proteins present in induced *E. coli* MC1061[F',pMSV$_{wt}$] cells (lane 1), or in induced *E. coli* MC1061[F',pMSV$_{Arg82-Ser}$] cells (lane 2) were analyzed by SDS-PAGE followed by staining with Coomassie Brilliant Blue R-250. The position of migration of gVp is indicated. An amount of protein equivalent to 0.3 ml of cell culture of OD_{600} = 1.8 (lightpath 1 cm) was loaded per lane.

DISCUSSION

A library of gene V mutants has been constructed by random chemical mutagenesis. Mutants were selected by a negative genetic complementation assay and their biological properties were measured. Furthermore, the level of gVp synthesis was established.

Negative genetic complementation, and translational repression assay

The biological properties of the mutant proteins have been studied by means of two different *in vivo* assays. The β-galactosidase assay only monitors the 'translational repressor activity' of the gVp mutants (Zaman *et al.*, 1990). However, the decreased production of phages and transducing particles by gVp containing cells, as observed in the negative genetic complementation assay, might be the result either of binding of gVp to the penetrating viral genome, which would inhibit the formation of RFI DNA, or of translational repression by gVp of the synthesis of gIIp, which would prevent initiation of rolling-circle replication from parental RFI DNA. To distinguish between these possibilities, a negative genetic complementation assay was performed in which induced and non-induced *E. coli* MC1061[F'] cells harbouring phagemid pMSV$_{wt}$ were infected either with helper phage R408 or with a derivative of R408 containing a deletion of 14 nucleotides corresponding to nucleotides 2 to 15 of gene II mRNA (G.J.R. Zaman and R.N.H. Konings, unpublished results). It has been shown that introduction of this deletion in the II'/'*lacZ* reporter gene eliminates the gVp mediated translational repression (Michel and Zinder 1989a; Zaman, 1991). Upon superinfection with phage R408, the production of transducing particles by induced cells was 407 times less than that of non-induced cells (Table 2.2). When instead of R408, the deletion mutant was used, the production of transducing particles by induced cells was only 80 times less than that of non-induced cells. This observation indicates that the interference of gVp with the propagation of the helper phage and the phagemid DNA transducing particles is caused by both inhibition of the conversion of the infecting viral strand into RFI DNA and repression of the synthesis of gIIp. A similar conclusion was reached by others, based on transfection (Fulford and Model, 1988a), and complementary negative genetic selection experiments (Terwilliger *et al.*, 1988).

Formation of 'inclusion bodies'

Protein folding *in vivo* is a complex process. The outcome of this process is influenced by many factors, such as cofactors, prosthetic groups, chaperones and the overall intracellular physiological environment, including temperature and ionic strength (Mitraki and King, 1989). GVp mutations leading to charge loss, charge gain, charge decrease or reversal, or leading to polarity reversal, nearly always cause the mutant protein to accumulate in the cells as insoluble aggregates (Table 2.2), indicating that minor changes in the electrostatic, or hydrophobic interactions cause the polypeptide chain to fold differently, thus allowing insoluble aggregates to be formed (Mitraki and King, 1989; Fane and King, 1991). This is clearly illustrated by mutant proteins Ile2-Val *vs.* Ile2-Thr, and Ile6-Val *vs.* Ile6-Thr. Substitution of isoleucine by another apolar residue of the same size does not result in insoluble aggregates, but substituting it by the polar threonine does lead to insoluble aggregates. Besides changes in electrostatic, and hydrophobic interaction, steric hindrance may also be of importance. For instance, although mutant proteins Leu49-Phe and Ala55-Val do not have a change in charge or polarity, they accumulate in 'inclusion bodies' (Table 2.2). Proper folding of these mutants therefore must be impaired by the fact that there is not enough space in the native protein to accomodate the substituted residue.

Residues involved in ssDNA binding and protein-protein interactions

^{1}H-NMR studies have revealed that only two aromatic residues, Tyr26 and Phe73, and one aliphatic residue, Leu28, are involved in binding to short

stretches of ssDNA (Chapter 4; King and Coleman, 1987). Near to the nucleotide interaction site formed by these three residues lies a phosphate-binding electropositive cluster formed by Arg16 and Arg21, and Lys24 and Lys46 (King and Coleman, 1987). Furthermore, Tyr41 is involved in dimer-dimer contact (King and Coleman, 1988; van Duynhoven *et al.*, 1990; Folkers *et al.*, 1991). Based on the crystal structure, Phe68 is believed to be involved in the stabilization by hydrophobic interactions of the monomer-monomer contact in the dimer (Brayer and McPherson, 1983).

Seven mutants (*i.e.* Arg16-Cys, Arg16-His, Tyr26-Cyc, Tyr41-His, Tyr41-Phe, Phe68-Leu, Phe73-Leu) were obtained at five sites that are involved in ssDNA binding or protein-protein interaction. All of them lost both their ability to impair the propagation of the helper phage and their ability to repress the expression of the II'/'*lacZ* reporter gene at the level of translation. Mutants Arg16-His and Arg16-Cys lost the positive charge which is needed to interact with the phosphate-backbone of the nucleic acid. Mutants Tyr26-Cys and Phe73-Leu lost the aromatic side chain that is necessary to stack on the bases of the nucleic acid. A change in the hydrophobicity of Phe68-Leu may cause the dimer to be less stable and therefore reduce the biological functions of gVp. Is is not yet clear in what way Tyr41 is involved in dimer-dimer interaction. Mutant proteins Tyr41-His and Tyr41-Phe indicate that not just an aromatic side chain is required at position 41 but rather a hydrophilic residue.

Classification of mutant proteins

The mutants can be subdivided in different classes both on the basis of their 'packaging inhibition activity' and their 'translational repressor activity'. For convenience, a 'packaging inhibition activity' of 40 or higher was not considered as a substantial reduction of the inhibition of the packaging. Accordingly, a 'translational repressor activity' of 10 or higher was not considered as a substantial reduction of the translational repressor function.

Following these criteria, it can be deduced that most gVp mutants exhibited a similar behaviour in both assays: most mutants that are severely impaired in 'translational repressor activity' are also disabled in the inhibition of phagemid production. Similarly, most mutants that efficiently inhibit phagemid production, also efficiently repress the translation of the II'/'*lacZ* reporter gene (*e.g.* Pro8-Thr, Ile6-Val, and Thr14-Ala). There are, however, a few clear exceptions. Firstly, five mutants that are severely impaired in translational repressor function, still efficiently inhibit phagemid production (*i.e.* Glu5-Lys, Asp36-Asn, Asn39-Ile, Ile47-Thr, and Pro58-Ser). On the other hand, two mutants have a weak inhibitory effect on the production of the fusion protein than on the production of transducing particles (Ile6-Thr, and Lys7-Arg).

The observation that these mutants behave differently in the two employed *in vivo* assays, demonstrates that different gVp residues are involved in the binding to ssDNA and in the binding to the 5' non-translated leader of gene II mRNA. This suggests that gVp comprises, besides the domain that is involved in the aspecific nucleic acid binding, additional residues required for specific binding of gVp to the 5' leader sequence of gene II mRNA.

CHAPTER 3

FLUORESCENCE STUDIES OF THE BINDING OF BACTERIOPHAGE M13 GENE V MUTANT PROTEINS TO POLYNUCLEOTIDES

A.P.M. STASSEN, B.J.M. HARMSEN, J.G.G. SCHOENMAKERS, C.W. HILBERS, and R.N.H. KONINGS (1992) *Eur. J. Biochem.* **206**: 605-612.

ABSTRACT

This investigation describes how the binding characteristics of the single-stranded DNA binding protein encoded by gene V of bacteriophage M13, are affected by single-site amino acid substitutions. The series of mutant proteins tested, includes mutations in the purported monomer-monomer interaction region as well as mutations in the DNA binding domain at positions which are thought to be functionally involved in monomer-monomer interaction or single-stranded DNA binding. The characteristics of the binding of the mutant proteins to the homopolynucleotides poly(dA), poly(dU), and poly(dT), were studied by means of fluorescence titration experiments. The binding stoichiometry and fluorescence quenching of the mutant proteins are equal to, or lower than, the wild-type gene V protein values. In addition, all the proteins measured bind in a more or less co-operative manner to single-stranded DNA. The binding affinities for poly(dA) decrease in the following order: Tyr61-His > wild-type > Phe68-Leu and Arg16-His > Tyr41-Phe and Tyr41-His > Phe73-Leu > Arg21-Cys > Tyr34-His > Gly18-Asp/Tyr56-His. Possible explanations for the observed differences are discussed. The conservation of binding affinity, also for mutations in the single-stranded DNA binding domain, suggests that the binding to homopolynucleotides is largely non-specific.

INTRODUCTION

The gene V encoded protein (gVp) of the F-specific *Escherichia coli* bacteriophage Ff (M13, fd, f1) is a relatively small (87 amino acids) single-stranded DNA (ssDNA) binding protein. In the life cycle of the phage it fulfils at least two essential functions. Firstly, gVp switches late in infection when its concentration has reached a certain critical threshold level, the replication of the phage genome from replicative form formation to progeny ssDNA formation by binding in a co-operative way to the ssDNA of the phage (Alberts *et al.*, 1972; Pretorius *et al.*, 1975; Pörschke and Rauh, 1983; for a review, see Model and Russel, 1988). Secondly, it represses, among other phage genes (Zaman *et al.*, 1991), the translation of gene II via specific binding to the 5' non-translated leader of gene II mRNA (Michel and Zinder, 1989a; 1989b; Zaman *et al.*, 1990).

Pretorius *et al.* (1975) showed that the fluorescence of gVp is quenched upon binding to ssDNA. Making use of this quenching effect, the binding parameters of various protein•polynucleotide complexes were determined quantitatively (Alma *et al.*, 1983b; Bulsink *et al.*, 1985). Titration experiments with different homopolynucleotides showed a large variation in binding affinity (Bulsink *et al.*, 1985). The binding is to a large extent caused by electrostatic interactions and in all cases a stoichiometry (*n*) of approximately 4 nucleotides per protein monomer has been found (Bulsink *et al.*, 1985). In contrast, fluorescence depolarisation studies (Bulsink *et al.*, 1986), circular dichroism measurements (Kansy *et al.*, 1986) and ^{1}H-NMR studies (Alma *et al.*, 1982) have revealed that on binding to oligonucleotides, the number of nucleotides per protein monomer is 3.

The structure of gVp has been determined crystallographically by Brayer and McPherson (1983). They also deduced a model for the gVp•ssDNA complex (Brayer and McPherson, 1984a; 1984b). According to this model two DNA strands bind in an antiparallel fashion to one dimer with a stoichiometry of 5 nucleotides per protein monomer. The aromatic residues Tyr26, Tyr34, and Tyr41 of one monomer and Phe73 of the opposite one in the dimer, are involved in the binding by stacking with the bases of the nucleic acid. The residues Arg16, Arg21, Arg80 and Lys46 were believed to be involved in the binding by ionic

interactions with the phosphate moieties of the nucleic acid.

In solution the structure of gVp and of its complex with oligonucleotides has been studied extensively by means of ^{1}H-NMR techniques (Chapter 4; Alma *et al.*, 1981a; 1981b; 1982; 1983a; King and Coleman, 1987; 1988; van Duynhoven *et al.*, 1990; Folkers *et al.*, 1991). These experiments have shown that the structure of these complexes is different from the structure proposed by Brayer and McPherson. ^{1}H-NMR binding studies of wild-type gVp with oligonucleotides containing the covalently attached spin-label 4-hydroxy-1-oxyl-2,2,6,6-tetramethyl piperidine (TEMPO), have been used to locate the DNA binding domains in the protein (de Jong *et al.*, 1989b; van Duynhoven *et al.*, 1993; Folkers *et al*, 1993b). A spin-labelled ligand bound to a macromolecule selectively broadens the resonance signals of spin-systems of this macromolecule, which are close enough to the spin-label in the complex. This means that spin-label induced relaxation can be expected for nuclei which are situated in a DNA binding domain of the protein In this way it has been shown that in ssDNA binding two binding domains of gVp are involved (van Duynhoven *et al.*, 1993; Folkers *et al.*, 1993b). The first domain comprises residues Arg16 through Glu30 and residues Leu44 through Asp50, and the second domain Lys69 through Arg80. The residues of which the signals are broadened most and therefore must be in the most close vicinity to the oligonucleotide, are Gly18, Ser20, Tyr26, Leu28, and Phe73. A smaller effect was observed for resonances arising from Arg16, Ser17, Val19, Arg21, Gln22, Lys24, Ser27, Glu30, Lys46, Thr48, Asp50, Lys69, Gln72, Gly74, Ser75, Asp79, and Arg80 which implies that these residues are more distant from the spin-label. In addition, spin-labelled oligonucleotide binding studies with gVp missense mutants revealed that Tyr26 and Phe73 are the only two aromatic residues that are involved in binding (Chapter 4). These results are supported by earlier studies, which have indicated that in ssDNA binding, an electropositive phosphate-binding cluster is involved consisting of the aromatic residues Tyr26 and Phe73,, Leu28, and the positively charged residues Arg16, Arg21, Lys24, and Lys46 (King and Coleman, 1987).

Besides what is known about the amino acid residues involved in ssDNA binding, there is strong evidence that Tyr41 is involved in dimer-dimer interaction (Chapter 4, King and Coleman, 1988). Furthermore, the involvement of Phe68 in the monomer-monomer contact was deduced from the crystal structure (Brayer and McPherson, 1983).

Considering both what is known about the involvement of certain residues in the binding of gVp to ssDNA, and the fact that electrostatic interactions from arginine or lysine, as well as stacking interactions from tyrosine or phenylalanine are involved in the binding, the question arises how the binding characteristics of gVp are affected by missense mutations of these amino acids at the above mentioned positions In this respect Arg16, Arg21, Tyr26, and Phe73 are most interesting, because they show a fairly large line-broadening in the binding studies with spin-label.

The missense mutants of this study were taken from a library of mutants of M13 gene V (Chapter 2). To obtain this library, gene V, carried on a phagemid expression vector in which gene V is under control of an inducible promoter, was subjected to random chemical mutagenesis. The mutant proteins were characterized both with respect to their ability to inhibit the production of phagemid DNA transducing particles and their ability to repress the translation of an mRNA transcribed from a chimeric *lacZ* reporter gene whose expression is controlled by the promoter and translational initiation signals of M13 gene II. For this reason the substitutions were not predetermined.

Fortunately, the mutant proteins Arg16-His, Arg21-Cys, Tyr26-Cys, and Phe73-Leu were available for the investigation of the role of Arg16, Arg21, Tyr26, and Phe73 in the binding process. According to the crystal structure the domain responsible for the monomer-monomer interaction comprises approximately the residues Leu60 through Ala85 (Brayer and McPherson, 1983). Thus missense mutations in this region are also interesting to study a possible effect of the monomer-dimer equilibrium on the binding. Available were Tyr61-His, Phe68-Leu, Phe73-Leu, and Arg82-Ser. The importance of Tyr41 for the dimer-dimer interaction and its involvement in the co-operativity of the binding leading to large clusters of protein along the nucleotide lattice, was studied by the two missense mutants Tyr41-Phe, and Tyr41-His. In addition, also the effect of mutations on the positions of the two remaining tyrosines Tyr34 and Tyr56 were investigated using Tyr34-His, and Gly18-Asp/Tyr56-His. The latter mutant protein was only available as a double mutation.

In this paper we describe the binding characteristics of these mutant proteins as derived from fluorescence titration experiments for the complexation with poly(dA), poly(dU), and poly(dT).

MATERIALS AND METHODS

Materials

The procedure used for the construction and expression of a library of M13 gVp mutants has been described in Chapter 2. Regarding the level of expression, the mutants can be subdivided into two categories. The mutants of the first category reached a level of expression comparable to M13 infected cells. These proteins (Tyr41-Phe, Tyr41-His, and Phe68-Leu) are soluble and were isolated by the standard isolation procedure as described previously (Garssen *et al.*, 1977; van Duynhoven *et al*, 1990). On the other hand, the mutants of the second category reached a level of expression that is at least 10 times higher than the level of expression in wild-type M13 infected cells These proteins (Tyr26-Cys, Tyr34-His, Gly18-Asp/Tyr56-His, Tyr61-His, Phe73-Leu, Arg16-His, Arg21-Cys and Arg82-Ser) are to a large extent insoluble and are deposited in the cells in so-called 'inclusion bodies'. Because of this reduced solubility the standard isolation procedure had to be modified. After sonification of 20 g cells and subsequent centrifugation, the pellet

with the insoluble mutant protein, was treated for 4-15 hr at 4 °C, with 250 ml of a buffer containing 6 M guanidine hydrochloride, 0.02 M Tris• HCl (pH 7.8), 5 mM EDTA, 1 mM 2-mercaptoethanol, 0.05 M NaCl, and 10 % (vol./vol.) glycerol, resulting in the dissolution of the mutant gVp. After centrifugation the guanidine hydrochloride was removed from the supernatant by dialysis during 16 hr at 4°C against 3 x 4 l of the same buffer without guanidine hydrochloride. During this procedure a small amount of protein was precipitated, which was removed by centrifugation. Subsequently, the gVp was isolated as described previously for wild-type gVp (Garssen *et al.*, 1977; van Duynhoven *et al.*, 1990). The Phe73-Leu mutant protein could be isolated from the soluble as well as the insoluble fraction of the cells.

The concentrations of wild-type and mutant gVps were determined from ultraviolet-absorption measurements at 276 nm. Since the absorbance at this wavelength arises mainly from the five tyrosyl residues in the protein, especially replacement of one of these residues by another amino acid will drastically decrease the value of the molar absorption coefficient in the order of magnitude of 20%. Therefore these coefficients were determined for the mutant proteins by means of the Bio-Rad Protein Assay (Bradford, 1976) using wild-type gVp as a reference. The following molar absorption coefficients were used: 6400 M^{-1} cm^{-1} for Tyr26-Cys, 5150 M^{-1} cm^{-1} for Tyr34-His, 5680 M^{-1} cm^{-1} for Tyr41-Phe, Tyr41-His, and Tyr61-His, 5370 M^{-1} cm^{-1} for Gly18-Asp/Tyr56-His, 7100 M^{-1} cm^{-1} for Phe68-Leu, Arg16-His, Arg21-Cys, Arg82-Ser, and wild-type gVp, and 8000 M^{-1} cm^{-1} for Phe73-Leu. The polynucleotides (poly(dA), poly(dU) and poly(dT)) with an average length of 150 nucleotides, were purchased from Pharmacia (Uppsala, Sweden). Their concentrations were derived from ultraviolet-absorption measurements using, as molar absorption coefficients, 9300 M^{-1} cm^{-1} (260 nm) for poly(dA), 9200 M^{-1} cm^{-1} (260 nm) for poly(dU), 8700 M^{-1} cm^{-1} (265 nm) for poly(dT) (Bulsink *et al.*, 1985). All ultraviolet-absorption measurements were performed with a Zeiss PMQII spectrophotometer.

Fluorescence titrations

Fluorescence measurements were performed with a Perkin Elmer fluorescence spectrophotometer MPF-4 equipped with a 150 W Xenon lamp. The instrument was used in the ratio mode which means that relative fluorescence intensities were measured instead of absolute quantum yields. Excitation was at the absorption maximum (276 nm) of the tyrosyl residues and the fluorescence was measured at the emission maximum (303 nm). All measurements were performed at 5 °C with a 5 mm absorption pathway. The fluorescence intensities, measured in arbitrary units, were corrected for Raman-scattering, the inner filter effect, *i.e.* correction for the loss of excitation intensity due to absorption by the dissolved protein in particular, and dilution during titration. For the inner filter effect correction, the factor $x/(1-e^{-x})$ was used, in which $x = 0.5 \cdot \ln 10 \cdot (\varepsilon_P \cdot c_P + \varepsilon_N \cdot c_N)$, and ε_P and ε_N are the molar absorption coefficients at 276 nm of protein and polynucleotide, respectively, and c_P and c_N are the protein and polynucleotide concentrations, respectively (see Lakowicz, 1983).

Two types of fluorescence titration experiments were performed:

(1) To measure binding isotherms small volumes of a protein solution, containing 10 mM sodium cacodylate (pH 6.9) and the desired sodium chloride concentration, were added to a polynucleotide solution, containing also 10 mM sodium cacodylate (pH 6.9) and the same sodium chloride concentration.

(2) In experiments where the dissociation of the protein•polynucleotide complex was studied, a concentrated solution of sodium chloride, containing 10 mM sodium cacodylate (pH 6.9), was added to a solution (with 10 mM sodium cacodylate (pH 6.9) and little salt) of the complex in the presence of a slight excess of polynucleotide, ensuring that all the protein is bound.

The binding parameters were deduced from the titration curves according to the methods described previously (Alma *et al.*, 1983b; Bulsink *et al.*, 1985).

RESULTS

The stoichiometry and fluorescence quenching

The characterization of the binding requires first of all the determination of the binding stoichiometry (n), *i.e.* the number of nucleotide residues covered by one protein monomer, and the fraction by which the fluorescence of the protein is quenched upon binding (Q_{max}). These two parameters are derived from binding isotherms at low salt conditions, when the intrinsic binding constant (K_{int}) is high, resulting in almost complete binding of the added protein until the polynucleotide is fully saturated (Alma *et al.*, 1983b). From the intersection of the two straight lines thus obtained the value of n is derived, while the ratio of the slopes of the two lines gives Q_{max}. Therefore the binding isotherms were determined for wild-type and mutant proteins at sodium chloride concentrations of 50 mM or even lower for the weaker binding proteins. Three polynucleotides were used, poly(dA), poly(dU), and poly(dT). An example of this type of titration with wild-type gVp is given in Figure 3.1A. This figure also shows the influence of the inner filter effect on the titration curve, which among other things results in a small deviation from linearity after the intersection point. Similar curves were obtained for the mutant proteins. The n and Q_{max} values found for the binding of the proteins to the polynucleotides are listed in Table 3.1. For wild-type gVp the number of nucleotides covered by one protein monomer turned out to be 3.6-3.8 and the fluorescence quenching factor was 0.62-0.66 (without the inner filter effect correction, n changes from 3.6 to 3.9, and Q_{max} from 0.62 to 0.58). These values are consistent with those reported in earlier publications (Alma *et al.*, 1983b; Bulsink *et al.*, 1985). Both for wild-type and mutant gVps, n and Q_{max} are not dependent on the salt concentration in the range of interest (data not shown: see also Discussion).

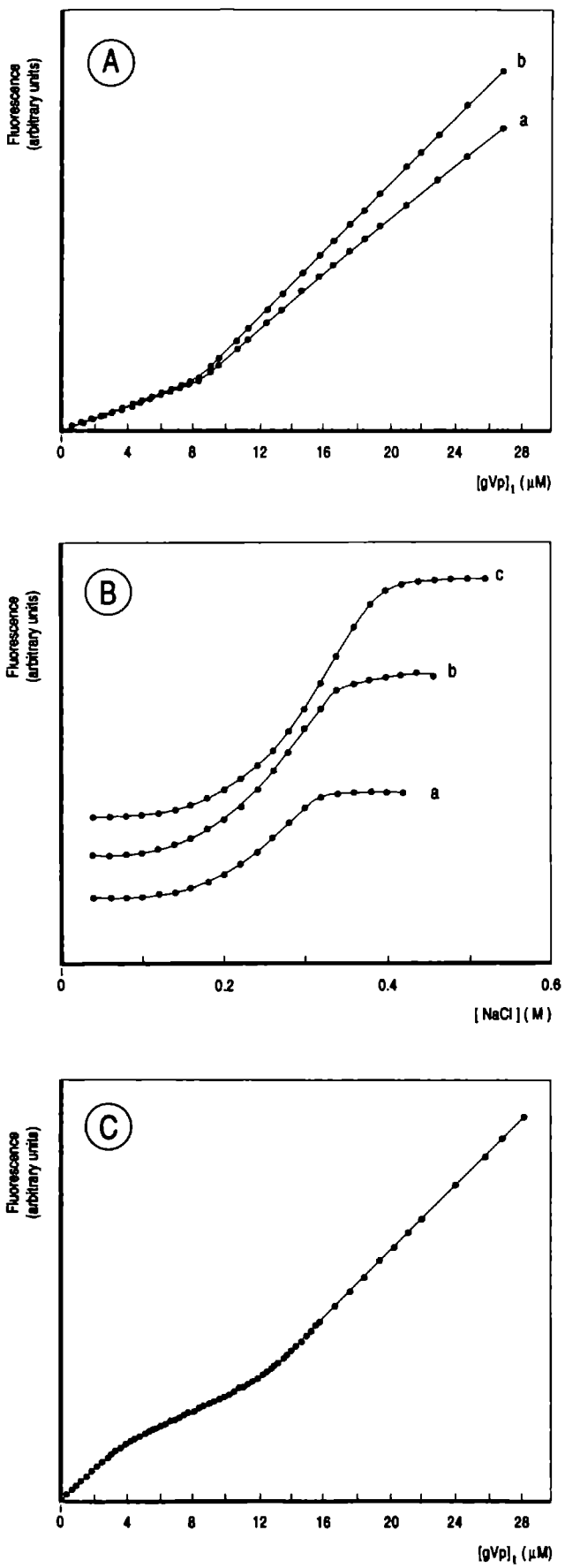


Figure 3.1 Fluorescence titration experiments of wild-type gVp and poly(dA).
(A) Binding isotherm at low (50 mM) sodium chloride concentration, before (a) and after (b) correction for the inner filter effect.
(B) gVp fluorescence as a function of the sodium chloride concentration upon disruption of the gVp•poly(dA) complex. The gVp concentrations are 12 μM (a), 20 μM (b), and 27 μM (c).
(C) Binding isotherm at high (200 mM) sodium chloride concentration.
All titrations were performed in 10 mM sodium cacodylate (pH 6.9). $[gVp]_t$ is the total gVp concentration.

The n and Q_{max} values of many mutant proteins deviate, however, considerably from the wild-type values. In this respect it is remarkable that the Q_{max}/n ratios do not differ significantly from the wild-type value for most of the mutant proteins. Possible explanations for these observations will be given in the Discussion.

The effective binding constant

The effective binding constant ($K_{int} \cdot \omega$, in which ω is the co-operativity parameter) was derived from titration curves representing the dissociation of the protein•polynucleotide complex upon addition of NaCl, as well as from binding isotherms at higher ionic strength at which the co-operative binding behaviour and the lower binding affinity lead to titration curves with a sigmoidal shape (Alma *et al.*, 1983b; Bulsink *et al.*, 1985). The latter type of titration also permits the determination of K_{int} and ω separately by means of a fitting procedure (*vide infra*) (Alma *et al.*, 1983b).

A salt titration experiment is started at low ionic strength (about 10 mM NaCl) at which the polynucleotides are fully saturated, while a slight excess of polynucleotide ensures that there is very little free protein in the solution. It is noted that the binding stoichiometry, deduced from the binding isotherm at low ionic strength, is used for the calculation of the concentrations leading to full saturation. Upon addition of sodium chloride, the complex will dissociate, which is monitored by an increasing fluorescence intensity (see Figure 3.1B). From the free protein concentration at the inflection point of the curve (*i.e.* at half-saturation, see Bulsink *et al.*, 1985) the effective binding constant can be derived for the sodium chloride concentration at this point. By performing these experiments at various concentrations of the nucleic acid•protein complexes, the effective binding constant can be determined as a function of the sodium chloride concentration.

The salt dependencies of the effective binding constant ($K_{int} \cdot \omega$) for wild-type and mutant gVp•poly(dA) complexes are presented in Figure 3.2, where $\log(K_{int} \cdot \omega)$ is plotted as a function of log[NaCl]. According to the theory of Record *et al.* (1976; 1978) such plots should show a linear relationship with:

$$\frac{-d\log(K_{int} \cdot \omega)}{d\log[NaCl]} = m'\psi + k,$$

where $m'\psi$ is the number of Na^+ ions released from the

Table 3.1 Binding characteristics at 5 °C and pH 6.9, deduced from binding isotherms at low [NaCl], for the binding of wild-type and mutant gVps to poly(dA), poly(dU), and poly(dT).

gVp	poly(dA)			poly(dU)			poly(dT)		
	Q_{max}	n	Q_{max}/n	Q_{max}	n	Q_{max}/n	Q_{max}	n	Q_{max}/n
wild-type	0.62	3.6	0.17	0.62	3.8	0.16	0.66	3.7	0.18
Tyr26-Cys	0.50	2.0	0.25				0.52	2.2	0.24
Tyr34-His	0.49	3.1	0.16						
Tyr41-Phe	0.37	3.5	0.11						
Tyr41-His	0.22	1.9	0.12	0.34	2.9	0.12	0.37	3.3	0.11
Gly18-Asp/ Tyr56-His	0.50	2.0	0.25	0.54	2.6	0.21			
Tyr61-His	0.63	3.7	0.17						
Phe68-Leu	0.60	3.1	0.19	0.53	2.9	0.18			
Phe73-Leu	0.57	2.3	0.25	0.61	3.6	0.17	0.62	3.9	0.16
Arg16-His	0.37	2.9	0.13				0.63	3.1	0.20
Arg21-Cys	0.22	1.7	0.13	0.17	2.3	0.07	0.36	2.3	0.16
Arg82-Ser	0.47	3.1	0.15	0.55	3.9	0.14			

DNA and k is the number of Cl^- ions released from the protein upon complex formation. As follows from Figure 3.2, within the experimental accuracy, straight lines are observed for all the proteins studied, suggesting that Record's theory is applicable to both the wild-type and the mutant proteins. Except for Tyr26-Cys and Arg82-Ser, the slopes of these lines are found to be similar, *i.e.*, $-d \log (K_{int} \cdot \omega)/d \log [NaCl] = 4.0 \pm 0.5$, indicating a close resemblance in binding behaviour (see Table 3.2). For the wild-type protein this value is identical to our earlier results (Alma *et al.*, 1983b; Bulsink *et al.*, 1985). The most striking fact is the variation in binding affinities for the different mutant proteins as represented by the values of $K_{int} \cdot \omega$ at the same NaCl concentration. From the plots of Figure 3.2 these values were determined at 0.1 M NaCl and also listed in Table 3.2. It is clear that the mutant proteins compared to the wild-type protein, show a similar or a weaker affinity for poly(dA). For Gly18-Asp/TyrY56-His, the only mutant protein with an amino acid substitution at two different sites, the binding affinity is reduced by a factor of more than 1000. The binding affinities for poly(dA) decrease in the following order: Tyr61-His > wild-type > Phe68-Leu and Arg16-His > Tyr41-Phe and Tyr41-His > Phe73-Leu > Arg21-Cys > Tyr34-His > Gly18-Asp/Tyr56-His.

The co-operativity parameter

The effective binding constant ($K_{int} \cdot \omega$) can be determined accurately from salt titration curves. However, from these experiments the separate contributions, *i.e.* the intrinsic binding constant and the co-operativity parameter, to the binding affinity cannot be derived. Therefore these parameters were determined separately from binding isotherms at not too low ionic strength. To obtain a reliable estimate of the magnitude of the co-operativity parameter it is essential that at the start of the titration, when the concentration of the added protein is low, nearly no binding occurs. This can be achieved by choosing the right NaCl concentration, because the binding affinity decreases with increasing salt concentration (*vide supra*). From the observed salt dependency of $K_{int} \cdot \omega$ (see Figure 3.2) the most adequate salt concentration can be easily calculated. Upon the first addition of protein, molecules bind with an affinity equal to K_{int}, because only isolated binding sites are available (see Figure 3.1C). Subsequently, binding to contiguous binding sites is preferred as a result of the enhanced binding affinity ($K_{int} \cdot \omega$) for this type of interaction. These effects are illustrated by the sigmoidal shape of the titration curve as shown in Figure 3.1C. This implies that the shape of the titration curve thus obtained is dependent on the magnitude of ω. In the fitting procedure first an arbitrary value for ω was chosen and the binding isotherm was calculated according to the theory of McGhee and von Hippel (1974), using the known values for n, Q_{max}, and $K_{int} \cdot \omega$. This binding isotherm was compared with the experimental one and then ω was varied until the best fit with the experimental curve was obtained. In some cases the fitting procedure started with a value for ω derived from the experimental data according to a procedure introduced by Watanabe and Schwarz (1983; Schwarz and Watanabe, 1983).

To obtain binding isotherms suitable to subject to the fitting procedure, sodium chloride concentrations in the range from 15 mM to 200 mM were chosen, depending on the effective binding constant of the particular protein. The co-operativity parameters determined in this way for the binding to poly(dA) are also listed in Table 3.2. The co-operativity parameter of wild-type gVp agrees with the value found previously (Alma *et al.*, 1983b; Bulsink *et al.*, 1985). Like wild-type gVp, all mutant proteins studied, bind co-operatively to

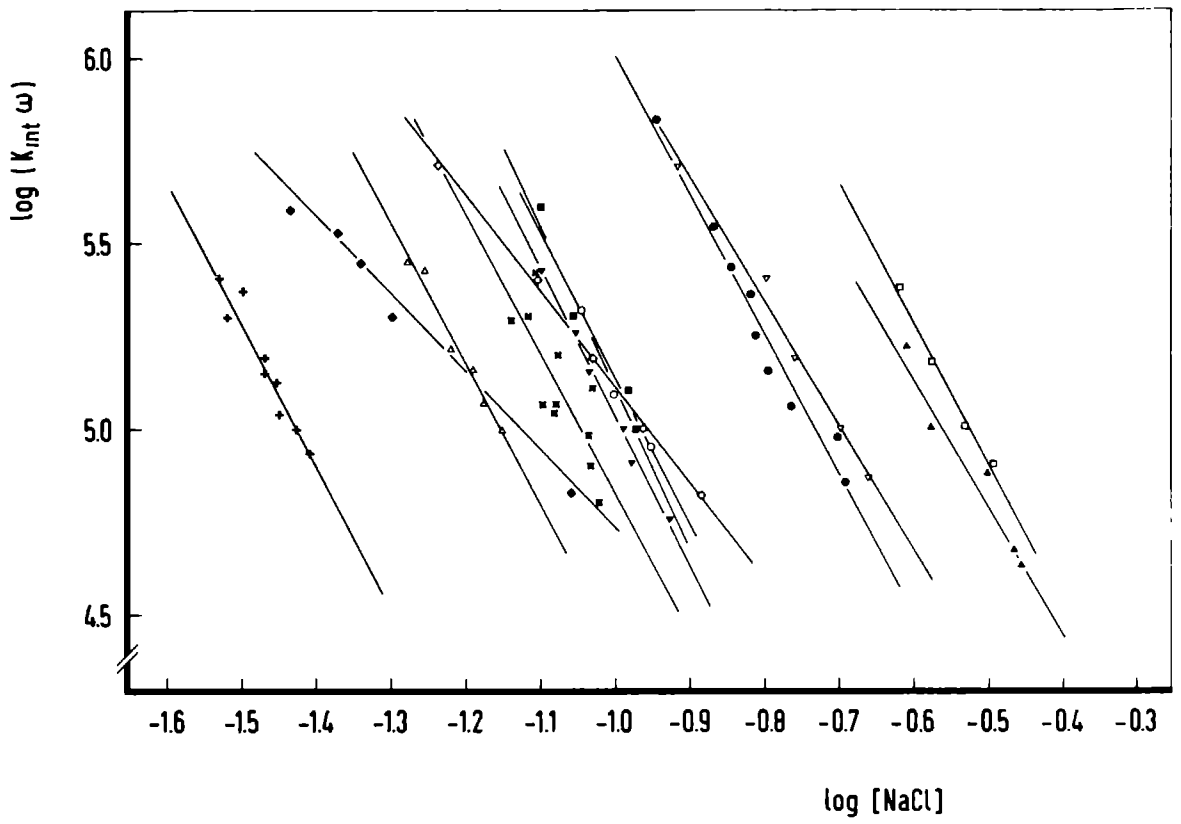


Figure 3.2 Salt dependency of the effective binding constant $K_{int} \cdot \omega$ for wild-type and mutant gVps, as derived from salt titrations of the complex containing poly(dA).
Gly18-Asp/Tyr56-His (+); Tyr34-His (Δ); Tyr26-Cys (◆); Arg21-Cys (x); Phe73-Leu (▼); Tyr41-His (■); Tyr41-Phe (O); Arg82-Ser (◊); Arg16-His (•); Phe68-Leu (∇); wild-type (▲); Tyr61-His (❑).
[NaCl] in M and $K_{int} \cdot \omega$ in M^{-1}.

poly(dA), although some of them, *e.g.* Tyr41-His, show a reduced co-operativity compared to the wild-type protein.

To determine binding isotherms suited for the calculation of the intrinsic binding constant and the co-operativity parameter, many datapoints are needed. Therefore the (mutant) protein has to be stable during a period up to 3 hr, needed to complete the experiment. The mutant proteins Tyr26-Cys and Tyr34-His turned out to be very unstable. For this reason such a binding isotherm could not be recorded for mutant Tyr34-His and only one with too few datapoints for Tyr26-Cys. However, from the shape of the isotherm of Tyr26-Cys it was concluded that Tyr26-Cys also binds co-operatively to poly(dA).

Effect of the solubilization procedure on the protein conformation

Because in a first instance many mutant proteins were obtained as an insoluble product, they had to be solubilized by treatment with guanidine hydrochloride followed by removal of this reagent via dialysis. In this respect the question is important as to whether this procedure alters the protein conformation permanently or not. Therefore a batch of wild-type gVp, isolated and purified according to the original standard procedure which does not involve the use of guanidine hydrochloride (Garssen *et al.*, 1977; van Duynhoven *et al.*, 1990), was subjected to treatment with guanidine hydrochloride and the dialysis procedure. After this, the binding properties (n, Q_{max}, ω, $K_{int} \cdot \omega$, and its salt dependency) were determined as usual from fluorescence titrations (data not shown) and compared with those found for the same batch before denaturation by guanidine hydrochloride. Within the experimental accuracy no differences could be observed, which indicates that, at least for wild-type gVp, the denaturation is completely reversible and that the protein can be simply renaturated by dialysis. That the denaturation-renaturation step does not lead to a protein with a different conformation was confirmed unambiguously by the ^{1}H-NMR spectra, which were identical, also with respect to details which are very sensitive to alterations in the protein conformation (Chapter 4).

As mentioned already (see Material and Methods) Phe73-Leu was isolated not only from an insoluble pellet by guanidine hydrochloride treatment, but also

Table 3.2 Binding characteristics at 5 °C and pH 6.9, for the binding of wild-type and mutant gVp to poly(dA), deduced from salt titration experiments and binding isotherms at high ionic strength.

gVp	$\frac{-d\log(K_{int} \cdot \omega)}{d\log[NaCl]}$	$K_{int} \cdot \omega$ at 0.1 M NaCl M^{-1}	ω
wild-type	3.5	$3.0 \cdot 10^6$	600
Tyr26-Cys	2.0	$5.1 \cdot 10^4$	
Tyr34-His	4.0	$2.6 \cdot 10^4$	
Tyr41-Phe	4.0	$1.3 \cdot 10^5$	220
Tyr41-His	4.5	$1.3 \cdot 10^5$	110
Gly18-Asp Tyr56-His	4.0	$2.2 \cdot 10^3$	185
Tyr61-His	4.0	$6.3 \cdot 10^6$	600
Phe68-Leu	3.5	$1.0 \cdot 10^6$	600
Phe73-Leu	4.0	$1.1 \cdot 10^5$	310
Arg16-His	3.5	$1.0 \cdot 10^6$	120
Arg21-Cys	4.0	$6.6 \cdot 10^4$	50
Arg82-Ser	2.5	$1.3 \cdot 10^5$	800

directly from the soluble fraction of the cell without guanidine hydrochloride. Again no differences in the binding characteristics of both samples were detectable and the ^{1}H-NMR spectra were also identical (Chapter 4). Therefore we conclude that the conformation and the properties of gene V proteins obtained after denaturation by guanidine hydrochloride and renaturation, are characteristic for these proteins in solution. This agrees very well with the results of Zabin and Terwilliger (1991). These authors found reversibility of guanidine hydrochloride-induced denaturation for wild-type gVp (Liang and Terwilliger, 1991) and for five temperature sensitive gVp variants, including Phe68-Leu, using circular dichroism measurements at 229 nm (Zabin and Terwilliger, 1991).

DISCUSSION

Stability of gene V mutant proteins

Fluorescence titration studies of wild-type gVp are complicated by the intrinsic instability of the protein, in particular at low protein concentrations and at low ionic strength, resulting in denaturation followed by the formation of higher protein aggregates (see *e.g.* Alma *et al.*, 1983b). For wild-type gVp artefacts due to denaturation can be avoided by making the appropriate precautions (Alma *et al.*, 1983b). However, all mutant proteins except Tyr41-His, Tyr41-Phe, and Phe73-Leu showed a further reduced stability under fluorescence titration conditions, as compared to the wild-type protein. The dependence of the protein denaturation on its concentration is clearly demonstrated by the different behaviour of some mutant proteins in ^{1}H-NMR experiments (protein concentration 1-5 mM; Chapter 4) and in fluorescence titrations (protein concentrations in the μM range). For example, Tyr26-Cys and Tyr34-His show the same, or even a higher stability with respect to the wild-type protein in ^{1}H-NMR measurements (Chapter 4), while fluorescence titrations carried out to determine K_{int} and ω separately, could not be performed with these proteins as a consequence of their instability at these concentrations (see Results). This concentration dependence can be explained by assuming that the denaturation proceeds via the following pathway:

(1) $D \leftrightarrow 2M$ equilibrium
(2) $M \rightarrow U$ denaturation
(3) $xU \rightarrow U_x$ aggregation

where D is the folded dimer, M is the folded monomer, and U is the unfolded monomer, which further aggregates to larger complexes U_x. We mention that in accordance with this model, from equilibrium sedimentation experiments with wild-type gVp an association constant for the monomer-dimer equilibrium (1) of $8 \cdot 10^5$ M^{-1} has been calculated (Pörschke and Rauh, 1983), which means that in ^{1}H-NMR measurements the protein is mainly present as a dimer, while in fluorescence titrations it is almost completely dissociated into monomers. A comparable model has been proposed by Liang and Terwilliger (1991) to describe the denaturation of gVp by guanidine hydrochloride. In that model the denaturation step 2 is reversible and does not lead to the formation of higher aggregates (the presence of aggregates is indeed very improbable at high guanidine hydrochloride concentrations). At guanidine hydrochloride concentrations above 2 M, partially folded intermediates and the folded monomer M are absent, and the denaturation equilibrium can be described by a simple two-state model. In our description the rate of the irreversible denaturation (or perhaps reversible denaturation followed by irreversible aggregation) is dependent on the concentration of the folded monomer M. According to the monomer-dimer equilibrium (1) a lowering of the total protein concentration increases the monomer fraction thus favouring denaturation.

In this way, a mutation could influence the denaturation in two different ways, namely by enhancing the rate of the denaturation step (a direct effect on the stability of the folded monomer), or by causing a shift of the monomer-dimer equilibrium to the monomer (increased dissociation constant). The concentration dependence of the protein stability provides an indication for which of these mechanisms is connected with a specific mutation. For example, as has been mentioned already, Tyr26-Cys and Tyr34-His are stable at high concentration (^{1}H-NMR conditions) but far less stable than wild-type gVp at low concentration (fluores-

cence titration conditions). This implies that the decreased stability is a property of the monomer itself. On the other hand, Tyr61-His and Phe68-Leu show a low stability at concentrations used for ^{1}H-NMR measurements (Chapter 4), which points to a strongly enhanced dimer dissociation at these protein concentrations, suggesting a direct involvement of these residues in the monomer-monomer interaction. For Phe68-Leu this result agrees with the observation of Zabin and Terwilliger (1991), who also suggest as a possible explanation for the strongly increased instability, the loss of the interaction between two Phe68 residues in a monomer-monomer contact upon replacement of these residues by leucine. It is not simple to predict the effect of a specific mutation on the protein stability. Generally it is assumed that, depending on the nature of the mutation (*e.g.* nonpolar *vs.* polar), substitutions at sites which are buried in the interior of the protein molecule, are more destabilizing than mutations at solvent exposed sites. However, it has been shown that out of ten temperature sensitive gVp mutants differing in stability from the wild-type protein, three mutations were at buried sites, four at moderate exposed sites, and three at highly exposed sites (Zabin and Terwilliger, 1991).

Binding properties of gVp with a mutation in the DNA binding domain

As mentioned already strong evidence exists that Arg16, Arg21, Tyr26, and Phe73 are directly involved in DNA interaction (Chapter 4; King and Coleman, 1987). The arginine mutants, Arg16-His and Arg21-Cys, exhibit significantly decreased values of n, Q_{max}, and ω. Caution is needed in the interpretation of the decreased n and Q_{max}, because these effects can originate from different sources. Substitution of an arginine by histidine, or cysteine does not necessarily affect the tyrosine fluorescence or its quenching Q_{max}, but it cannot be excluded either. Concerning the decrease of n, site sizes lower than 4 have also been found for the binding of M13 wild-type gVp to oligonucleotides, and to polynucleotides in the absence of salt, *i.e.* at salt concentrations lower than 5 mM (n = 2 or 3) (Alma *et al.*, 1982; Gray *et al.* 1984; Bulsink *et al.*, 1986; Kansy *et al.*, 1986; Bulsink *et al.*, 1988).

Comparison of the stoichiometric numbers of the different mutant proteins in Table 3.1, shows that n ranges from 1.7 to 3.7, rather than being either n=2, 3, or 4. In addition, a decrease in n is sometimes accompanied by a decrease in the fluorescence quenching Q_{max}, leading to the result that Q_{max}/n is less affected. For some of the mutant proteins, in particular for those with a low stability and an increased tendency to aggregate, the smaller values of n and Q_{max} may result from the presence of a considerable amount of the unfolded monomer U or its aggregates U_x in the solution, which does not bind to the polydeoxyribonucleotides. From an analysis of the fluorescence titrations it can easily be shown that the presence of inactive protein lowers the Q_{max} and n by the same factor, thus leaving Q_{max}/n unaffected. In this respect salt titration experiments are more reliable because they start with a solution of the protein• polynucleotide complex obtained by once mixing two freshly prepared solutions containing the protein and the polynucleotide in relatively high concentrations. Only Tyr26-Cys, Gly18-Asp/Tyr56-His, and Phe73-Leu show strongly increased Q_{max}/n values. Since the replacement of a tyrosyl residue will, in general, influence the tyrosyl fluorescence it is difficult to draw conclusions from changes of Q_{max} for this type of mutation. For Phe73-Leu, Q_{max} is close to the value observed for wild-type gVp. Here the most striking effect is the much lower n for the binding to poly(dA). It is remarkable that the binding to poly(dU) and poly(dT) yields the same value for n as for wild-type gVp. This mutant protein also has a lower binding affinity for poly(dA) as compared to the wild-type protein. Since binding to poly(dU) and poly(dT) is always stronger than to poly(dA), the decreased stoichiometry could be related to the lower affinity for poly(dA), maybe as a consequence of the loss of stacking interaction between Phe73 and an adenyl residue. This implies that the binding strength itself can affect stoichiometry or quenching. Similar observations have recently been reported by Zabin and Terwilliger (1991), who observed a decreased binding affinity accompanied with a lowering of n for the mutant proteins, Lys69-His and Lys69-Met. Possibly the binding affinity can change the stoichiometry by a modulation of the DNA structure.

Tyr26-Cys shows a binding behaviour completely different from that of wild-type gVp. A decrease of n to 2.0 is observed together with an increase of Q_{max}/n to 0.25, which is much higher than the corresponding value for wild-type gVp (Table 3.1). Therefore the lowering of n cannot be explained by the increased protein instability. Moreover, $K_{int} \cdot \omega$ is 60 times lower (Table 3.2). Furthermore, Tyr26-Cys and Arg82-Ser are the only mutant proteins for which the salt dependency of $K_{int} \cdot \omega$ is not the same as that of wild-type gVp. These observations are congruent with the current idea that besides Phe73, Tyr26 is also directly involved in interaction with the DNA (Chapter 4; King and Coleman, 1987).

For Gly18Asp/Tyr56-His, $K_{int} \cdot \omega$ is even 1400 times smaller than for the wild-type protein at the same NaCl concentration. Possibly the substitution of Gly18 is responsible for this lower binding affinity because Gly18 also is part of the DNA binding wing (van Duynhoven *et al.*, 1990; Folkers *et al.*, 1991).

Effect of mutations in the purported monomer-monomer interaction region on the binding properties

The interaction area of two monomeric subunits in the dimer is presumed to comprise roughly residues Leu60 through Ala86 (Brayer and McPherson, 1983). Therefore mutations in this sequence could result in a perturbation of the monomer-dimer equilibrium. For Tyr61-His and Phe68-Leu this was already demonstrated by the increased instability of these mutant proteins. This needs not automatically lead to a change of the DNA binding properties, because at the low

protein concentrations at which the fluorescence titrations are carried out the proteins probably are already present, mainly as the folded monomers M. If binding occurs from the monomeric state then a decrease of the dimer dissociation constant will have little influence on the binding isotherms measured under fluorescence titration conditions. The binding characteristics of Tyr61-His and Phe68-Leu are indeed nearly identical to those of wild-type gVp, indicating that both residues are not involved directly in the interaction with the DNA. For Phe68-Leu the results are in good agreement with the values recently reported by Zabin and Terwilliger (1991). The two other proteins with mutations in the purported monomer-monomer interactions area, *i.e.* Phe73-Leu and Arg82-Ser, show binding characteristics which are different from those of wild-type gVp. For Phe73-Leu this is not surprising because Phe73 is known to be in contact with the DNA upon binding. Arg82-Ser shows a drastically reduced salt dependency of the binding and also a lower affinity to poly(dA). The first effect can be explained by assuming that the replacement of a positively charged arginine at position 82 by a neutral residue results in the loss of a chloride ion bound at Arg82. For wild-type gVp this chloride ion is only released upon binding to ssDNA.

Effect of substitution at Tyr34 or Tyr41 on the DNA binding characteristics

In the original model of the protein•nucleic acid complex proposed by Brayer and McPherson (1983; 1984a) Tyr34 and Tyr41 are directly involved in the interaction with ssDNA. However, from ^{1}H-NMR studies (Chapter 4; King and Coleman, 1987) clear evidence has been obtained that Tyr26 and Phe73 are the only aromatic residues in direct contact with the nucleic acid in the DNA•protein complex. In this complex Tyr41 most likely plays a crucial role in the mutual interaction of gVp dimers. This may form the basis for the observed co-operativity in the binding to ssDNA and for the tendency of dimers to cluster to tetramers or even larger polymers in a protein solution at higher concentrations (Chapter 4). Indeed, substitution of Tyr41 leads to better solubility characteristics and better resolved ^{1}H-NMR spectra. The explanation for these effects can be found in a decrease of the tendency for clustering of the protein dimers after substitution of Tyr41 (Chapter 4).

Tyr34 might also be a part of the possible dimer-dimer interaction surface (Chapter 4), although this has not been proven yet. Tyr34-His has a 100 times lower binding affinity for poly(dA) than wild-type gVp. Since no values for ω could be calculated for this mutant protein from the binding isotherms, due to the instability of the protein, it is uncertain whether this smaller binding affinity results from a strong reduction of ω or a decrease of K_{int}, or of both. The two substitutions at position Tyr41 both show a reduction of ω with a factor 3 to 6 while $K_{int} \cdot \omega$ is about 20 times smaller than for wild-type gVp. Although for these mutant proteins a lower co-operativity parameter is indeed found, there is also an effect of the mutation on K_{int}. Moreover, as follows from Table 3.2, more mutant proteins have a reduced co-operativity. Because of the limited accuracy of the fitting procedure especially if inactive (denatured) protein is present, the separate values of ω have to be considered with caution.

For Tyr41-His the values of n and Q_{max} for the binding to poly(dA) are lower than those for wild-type gVp. However, higher values are found for poly(dU), while for poly(dT) n comes close to the wild-type value. This effect, namely the dependence of stoichiometry or quenching on the binding affinity itself, was also observed for Phe73-Leu (*vide supra*).

Relationship between biological activity and alterations in binding characteristics

In comparison to wild-type gVp, all mutant proteins considered in this study do not inhibit the production of phagemid DNA transducing particles, and are not able to repress the translation of a chimeric *lacZ* reporter gene, whose expression is controlled by the promoter and translational initiation signals of M13 gene II. It is, however, not simple to relate the *in vivo* behaviour of these proteins directly to their binding characteristics deduced from the present fluorescence studies. The interference of gVp with the propagation of the helper phage and the phagemid DNA transducing particles is caused by both inhibition of the conversion of the viral infecting strand into replicative form I DNA, in which non-specific nucleic acid binding is involved, and repression of the synthesis of gene II protein, which requires specific nucleic acid binding (Chapter 2). In addition, the production of insoluble protein in many cases may be a problem. In that situation the actual concentration of soluble, active protein in the cell is not known. A reduced biological activity could well be the result of the presence of only a very small amount of soluble protein instead of a change in binding properties. Substitutions at position Tyr41 do not have these drawbacks. This means that for Tyr41-Phe and Tyr41-His the reduction of the biological activity can be ascribed to their altered binding properties.

CONCLUSIONS

Single amino acid substitutions at crucial positions in the DNA binding domains of gVp lead to alterations in the binding characteristics without disturbing the binding completely, as far as the interaction with homopolydeoxyribonucleotides is concerned. The reason for this must be found in the fact that although there is a preference in binding affinity for certain bases, like U and T, the binding to homopolynucleotides is generally not very specific. In this respect it is now important for a better understanding of the biological significance of gVp•DNA interactions, to look for local details, *i.e.* to study the binding to certain essential DNA or RNA sequences. Recently, it has been shown that a 16-nucleotide-long region in the 5' non-translated leader of gene II mRNA, is required *in vivo* for repression by gVp (Michel and Zinder, 1989a; Zaman *et al.*, 1990; 1991). Mutations in this region which abolish gene II translational repression *in vivo*, also affect the binding of gVp to this region *in vitro*

(Michel and Zinder, 1989b; Zaman, 1991). The study by means of fluorescence titrations of the binding characteristics of wild-type and mutant gVp to these *in vivo* sequences is at present under investigation.

CHAPTER 4

IDENTIFICATION OF THE AROMATIC AMINO ACID RESIDUES OF THE GENE V PROTEIN OF BACTERIOPHAGE M13, INVOLVED IN SINGLE-STRANDED DNA BINDING

CHARACTERIZATION OF WILD-TYPE AND MUTANT M13 GENE V PROTEINS BY MEANS OF ^{1}H-NMR

P.J.M. FOLKERS, A.P.M. STASSEN, J.P.M. VAN DUYNHOVEN, B.J.M. HARMSEN, R.N.H. KONINGS, and C.W. HILBERS (1991) *Eur. J. Biochem.* **200**: 139-148.

ABSTRACT

Recording of good quality NMR spectra of the single-stranded DNA binding gene V protein of the bacteriophage M13 is hindered by a specific protein aggregation effect. Conditions are described for which NMR spectra of the protein can best be recorded. The aromatic part of the spectrum has been reinvestigated by means of two-dimensional total correlation spectroscopy. Sequence-specific assignments were obtained for all of the aromatic amino acid residues with the help of a series of single-site mutant proteins. The solution properties of the mutants of the aromatic amino acid residues have been fully investigated. It has been shown that, for these proteins, either none or only local changes occur compared to the wild-type molecule. Spin-labelled oligonucleotide-binding studies of wild-type and mutant gene V proteins indicate that Tyr26 and Phe73 are the only aromatic residues involved in binding to short stretches of single-stranded DNA. The degree of aggregation of wild-type gene V protein is dependent on both the total protein and salt concentration. The data obtained suggest the occurrence of specific protein-protein interactions between dimeric gene V protein molecules in which the tyrosine residue at position 41 is involved. This hypothesis is further strengthened by the observation that the solubility of Tyr41 mutants of gene V protein is significantly higher than that of the wild-type protein. The discovery of the so-called 'solubility' mutants of M13 gene V protein has finally made it possible to study the solution structure of gene V protein and its interaction with single-stranded DNA by means of two-dimensional NMR.

INTRODUCTION

One of the essential genes in the life cycle of filamentous bacteriophages, such as Ff (M13, fd, f1) and IKe, is gene V. The product of this gene is a single-stranded DNA binding protein (gVp) which is indispensable for the production of progeny viral strand DNA of the phage genome (Model and Russel, 1988). Late after infection, the protein binds strongly and co-operatively to single-stranded DNA (ssDNA) and initiates in this way the switch of double-stranded DNA synthesis to the production of progeny viral strands. Since it was first isolated, the protein has been subjected to various physico-chemical studies and is considered an important model for protein-ssDNA interactions. The protein consists of 87 amino acids, and exists in solution predominantly as a dimer (molecular mass 19.4 kDa). It is the only ssDNA binding protein whose crystal structure is currently known (Brayer and McPherson, 1983). The refined model of the protein reveals that the individual monomers are entirely composed of β-structure and that they are closely associated about a twofold axis. On the basis of the crystal structure, a model for the complex of ssDNA and gVp has been proposed (Brayer and McPherson, 1984a). In this interaction model, two nucleic acid strands bind to one dimer in an antiparallel fashion with stoichiometries of 5 nucleotides/protein monomer. The

proposed protein-ssDNA interactions involve electrostatic binding of the nucleic acid phospho-diester bond by residues Arg16, Arg21, Arg80 and Lys46 and stacking of the nucleic acid bases with the side chains of Tyr26, Tyr34 and Tyr41 of one monomer and Phe73 of the opposite monomer. In the model, the observed co-operativity in the binding of ssDNA to gVp has been explained in terms of an interaction between gVp dimers located on well defined domains of the individual protein molecules.

However, in previous papers, physico-chemical observations contrasting the proposed ssDNA-gVp interaction model, as well as of features of the crystal structure, have been reported (Gray *et al.*, 1984; King and Coleman, 1987, 1988; van Duynhoven *et al.*, 1990). One of the fundamental differences is the chosen stoichiometry of 5 nucleotides per monomer in the interaction model despite the fact that stoichiometries of 3 nucleotides per monomer for oligonucleotide•protein complexes (n=3 binding mode) and 4 nucleotides per monomer for polynucleotide•protein complexes (n=4 binding mode) have been reported (Alma *et al.*, 1981a; Bulsink *et al.*, 1986).

NMR studies have suggested the involvement of only two aromatic amino acid residues in the binding to ssDNA (King and Coleman, 1987). The assignments of these residues has been based on NOESY spectra recorded in D_2O solution combined with the knowledge of the crystal structure. However, ^{1}H-NMR assignments which are based on crystallographic data may easily lead to errors. This is particularly so given the fact that recently for the DNA binding wing in gVp, encompassing residues 13-31, a structure was proposed on the basis of ^{1}H-NMR data that differs from the same amino acid sequence in the crystal (van Duynhoven *et al.*, 1990). In order to obtain more information on the possible DNA binding domain and to establish unequivocally the roles of the individual aromatic amino acid residues in the complex formation with ssDNA, we started with the characterization of missense gVp mutants which have been constructed by means of saturation mutagenesis (Chapter 2).

This paper is the first in a series of NMR studies on M13 gVp. It describes general spectral and structural features of both wild-type (wt) gVp and a variety of gVp mutants. Special attention is paid to the gVp mutant proteins in which the aromatic residues are substituted for other residues. Solution properties of wild-type and mutant proteins, such as stability and solubility, are investigated. Optimal conditions for NMR experiments are established and the DNA binding domains of the proteins are explored by means of spin-labelled oligonucleotides. The results are compared with those obtained in earlier studies.

MATERIALS AND METHODS

Protein purification

The procedure used for the construction and expression of a library of M13 gVp mutants is described in Chapter 2. The level of expression of the mutant proteins can be divided into two categories. The low producers (wild-type, gVp Tyr41-His and Tyr41-Phe) reach a level of expression which is comparable to that in M13-infected cells and where the protein is present in the soluble fraction of the cells. The high producers (gVp Tyr26-Cys, Tyr34-His, Tyr61-His and Phe73-Leu) reach, for reasons still unknown, a level of expression which is at least ten times higher. These proteins accumulate in the form of highly refractile particles called 'inclusion bodies'. The soluble proteins were isolated as described previously (Garssen *et al.*, 1977; van Duynhoven *et al.*, 1990). A modified procedure was developed to isolate insoluble mutant proteins (*Cf.* Chapter 3). In short, after sonification and subsequent centrifugation of the cells, the pellet, which contained the insoluble mutant protein was resuspended in buffer containing 6 M guanidine hydrochloride, 0.02 M Tris•HCl pH 7.8, 5 mM EDTA, 1 mM 2-mercaptoethanol, 0.05 M NaCl, and 10% (vol./vol.) glycerol during 4-15 hr at 4°C. Then, the protein was renatured by means of dialysis in the same buffer without the guanidine hydrochloride. After subsequent centrifugation, the soluble protein fraction was loaded on a DNA-cellulose affinity column. From this point, the normal procedure for gVp isolation was reinstated. After dialysis against 1 mM cacodylate buffer (pH 6.9) containing 50 mM NaCl, all purified proteins were lyophilised and stored at -20°C. The gVp mutant Tyr41-His contains a histidine residue instead of a tyrosine residue at position 41. The other mutants are named in a similar manner.

The protein monomer concentrations were determined by measuring absorbance at 276 nm (pH 6.9) using the absorption coefficients of 6400 $M^{-1}cm^{-1}$ for Tyr26-Cys, 5150 $M^{-1}cm^{-1}$ for Tyr34-His, 5680 $M^{-1}cm^{-1}$ for Tyr41-Phe, Tyr41-His and Tyr61-His, 7100 $M^{-1}cm^{-1}$ for wild-type gVp, and 8000 $M^{-1}cm^{-1}$ for Phe73-Leu. The absorption coefficients were determined for the mutant proteins by means of the Bio-Rad Protein Assay (Bradford, 1976), using wild-type gVp as a reference.

Samples of wild-type and mutant gVp were measured in D_2O at very low ionic strength with a concentration ranging from approximately 0.5 to 4.0 mM dependent on the sample quality. The pH of the samples, which ranged between pH 5.1 and pH 7.5, was adjusted with diluted DCl.

Spin-labelled oligonucleotides

The spin-label trinucleotide, three adenyl residues to which the spin-label 4-hydroxy-1-oxyl-2,2,6,6-tetramethyl piperidine (TEMPO), was covalently attached via phospho-diester bonds at both the 5' and 3' end of the oligonucleotide, was synthesized, purified and characterized as described by Claesen *et al.* (1986). We will refer to this spin-label as: $*(dA)_3*$, where the asterisks refer to the attached spin-label. The concentration of $*d(A)_3*$ in H_2O (at pH = 7) was determined from its absorption at 260 nm, using the absorption coefficient for the adenine trimer (Cassani and Bollum, 1969) corrected for a slight extra absorption by the TEMPO moiety, *i.e.* 37500 $M^{-1}cm^{-1}$. The binding experiments were carried out by adding small amounts of a concentrated solution of the spin-labelled oligonucleotide to the protein solutions. The $*d(A)_3*$ (D_2O) solutions contained a low (millimolar) salt concentration and were

adjusted to the pH of the protein sample prior to the experiments.

^{1}H-NMR measurements

^{1}H-NMR experiments were performed at 400 MHz on a Bruker AM400 spectrometer interfaced to an ASPECT 3000 computer. MLEV17 TOCSY experiments (Bax and Davis, 1985; Griesinger *et al.*, 1988) were conducted at 298 K unless otherwise stated, with mixing times around 30 ms. The residual water resonance was suppressed using DANTE (Morris and Freeman, 1978). In all experiments the carrier was placed at the position of the water signal and time-proportional phase incrementation (TPPI) (Marion and Wüthrich, 1983) was used for signal accumulation in the t_1 dimension.

RESULTS

Optimization of solution conditions for ^{1}H-NMR analysis of M13 wt gVp

Conditions under which reasonable one-dimensional ^{1}H-NMR spectra of M13 wild-type gVp can be recorded were established previously (Alma *et al.*,1981a). Typically, the NMR samples contained 50 mM NaCl, 1 mM sodium cacodylate pH 7.0 and up to 1 mM gVp. However, we have discovered that it is possible to optimize the solution conditions further. Wild-type M13 gVp is particularly sensitive to alterations in protein concentration, salt concentration and pH. In Figure 4.1., changes which occur as a function of salt concentration are shown for the aromatic region of the ^{1}H-NMR spectrum of M13 wt gVp. The peaks belonging to a single tyrosine residue, which had been designated as TyrIV by Alma *et al.* (1981a), shift upfield when the salt concentration is increased. This effect is accompanied by a broadening of all spectral lines, indicative of an increase in aggregation. The increase of the protein concentration results in spectral alterations similar to those observed when the salt concentration is increased (data shown in: de Jong *et al.*, 1987b). Lowering the pH down to approximately 5.1 reverses the aggregation effect to some extent, while a further decrease results in rapid denaturation of the protein. Once denatured, it is not possible to renature the protein by increasing the pH. Hence, NMR measurements of M13 wt gVp can be performed best at very low ionic strength and at pH 5.1. Under these conditions, it was possible to study samples with concentrations up to 1.5 mM M13 wt gVp without inducing too significant aggregation and with reasonably narrow lines. The temperature range at which experiments can be conducted with these relatively concentrated samples is however very limited. Typically, NMR experiments can be performed at a temperature around 298 K.

Analysis of the aromatic region of the ^{1}H-NMR spectrum

The aromatic part of the ^{1}H-NMR spectrum of M13 wt gVp has been investigated thoroughly by Alma *et al.* (1981a). M13 wt gVp contains one histidine, three

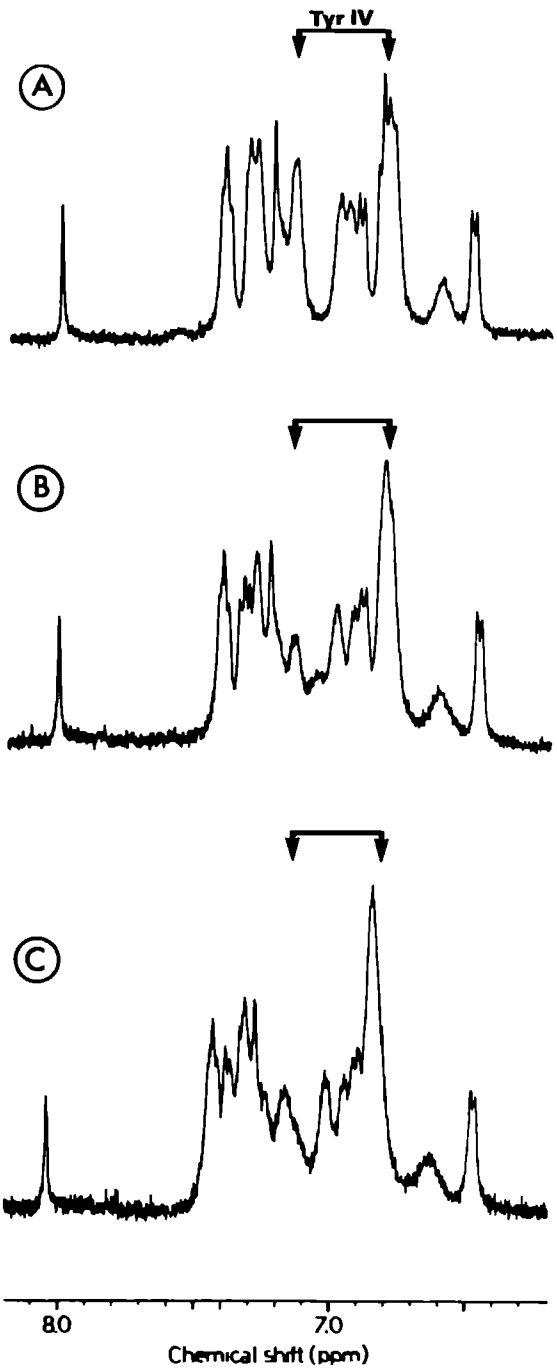


Figure 4.1 Salt concentration dependency of the aromatic region of the M13 wt gVp spectrum. The arrows indicate the resonance positions of the TyrIV spin-system. The sample contained 0.5 mM gVp, pH 7.5. Salt concentration: (A) low ionic strength; (B) 25 mM NaCl; (C) 100 mM NaCl.

phenylalanine and five tyrosine residues. The individual spin-systems of all aromatic residues have been identified. The unique histidine residue in gVp at position 64 could be sequence-specifically assigned. In Figure 4.2., a ^{1}H-NMR spectrum of the aromatic part of M13 wt gVp is presented in which the assignments of the individual protons to residue type are indicated by Roman numerals. The results were obtained by means of photo-chemically induced dynamic nuclear polarization experiments (Garssen *et al.*, 1978), selective decoupling experiments and one-dimensional NOE difference measurements (Alma *et al.*, 1981a).

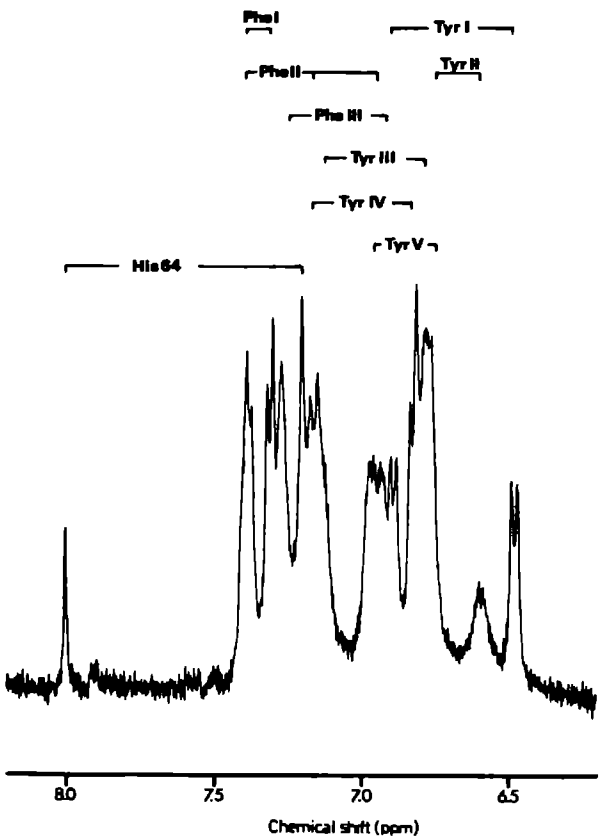


Figure 4.2 The aromatic part of the ^{1}H-NMR spectrum of 0.8 mM M13 wt gVp (pH 5.1).
The assignments of the individual protons to residue type are indicated by Roman numerals.

Nowadays, all coupled resonances in the aromatic region can be identified using two-dimensional double quantum filtered COSY (Rance *et al.*, 1983) and total correlation (TOCSY) spectra (Griesinger *et al.*, 1988). An example of a TOCSY spectrum of M13 wt gVp is given in Figure 4.3: Figure 4.3A displays a protein with a low degree of aggregation as can be judged from the titrations and pH titrations for this protein.The mutant proteins gVp Tyr26-Cys and Phe73-Leu behave very similarly in position of the TyrIV spin-system. In this situation, the TyrIV crosspeak is downfield shifted and is positioned left from the crosspeak connecting the resonances which make up the TyrIII spin-system. Figure 4.3B and C display a part of the aromatic region of two other TOCSY spectra of M13 wt gVp which reflect the influence of an increased aggregation. The crosspeak belonging to the TyrIV spin-system in Figure 4.3B is right on top of the crosspeak of spin-system TyrIII, while in Figure 4.3C it is even further upfield shifted. In the latter case, the resonances of spin-system TyrV are also slightly shifted. TyrII gives rise to broadened resonances indicative of intermediate exchange between the two possible chemical shift positions for the pairs of Hδ and Hε protons of the ring (Wüthrich, 1986). The remaining four tyrosines generate relatively narrow resonances characteristic of an aromatic ring flipping rapidly about the Cβ–Cγ bond such that the Hδ1 and Hδ2 resonances and the Hε1 and Hε2 resonances are superimposed. Similarly, the Phe resonances exhibit a degenerate spin-pattern due to rapid ring flips.

Characterization of mutants of the aromatic residues of M13 gVp

Missense mutants were constructed for four out of five tyrosines and one out of three phenylalanines. The level of gVp synthesis in the cells varies significantly with the mutants investigated. The level of expression of the wild-type protein and of the gVp mutants Tyr41-Phe and Tyr41-His is comparable to that of gene V protein in phage-M13-infected cells. These recombinant proteins are present in the soluble fraction and isolated as described previously (Garssen *et al.*, 1977). The expression level of the gVp mutants Tyr26-Cys, Tyr34-His, Tyr61-His and Phe73-Leu is, for reasons still unknown, at least ten times higher and their encoded proteins accumulate as so-called 'inclusion bodies' in the cell. The proteins were isolated from these particles using a denaturation step with guanidine hydrochloride. A significant amount of mutant protein Phe73-Leu, however, could also be isolated from the soluble fraction of the cell extract.

To test whether the renaturation procedure resulted in a proper folding of the polypeptide, ^{1}H-NMR spectra of wild-type M13 gVp isolated from the cytoplasmatic fraction and wild-type gVp that was subjected to the denaturation and renaturation procedure but derived from the same protein batch were compared (Figure 4.4). It can be concluded that these spectra are very similar, thus indicating that the renaturation procedure developed results in a proper refolding of the denatured protein. Close examination of the aromatic regions of the spectra, however, revealed a small difference in aggregation state between both samples, with the control, wild-type, sample being less aggregated.

Optimal samples of all mutant proteins were obtained at very low ionic strength, a property which is similar to wt gVp. The differences in quality of the mutant gVp samples varied significantly however. For example, gVp mutant Tyr61-His is very unstable. Samples of Tyr61-His could only contain up to 0.5 mM

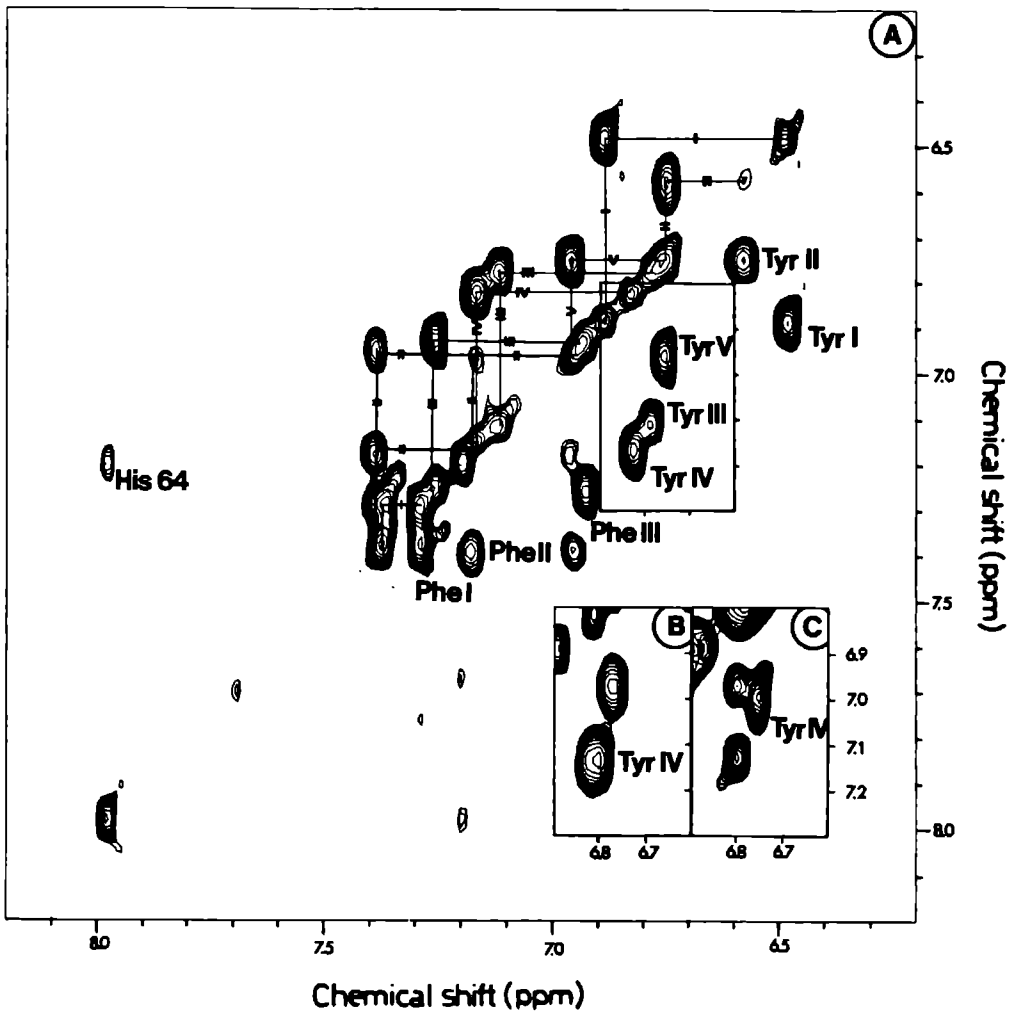


Figure 4.3 TOCSY spectrum of M13 wt gVp.
(A) 30 ms TOCSY spectrum of the aromatic region of 0.5 mM M13 wt gVp at low ionic strength (pH 7.1). The individual spin-systems are indicated by connecting arrows. The assignments to residue type are taken from Alma *et al.* (1981a). (B) TOCSY spectrum of a part of the aromatic region of wt gVp (1 mM, pH 7.0) at low ionic strength. (C) TOCSY spectrum of a part of the aromatic region of wt gVp at higher ionic strength (1 mM, pH 7.5), displaying upfield-shifted TyrIV resonances.

protein and experiments could only be conducted at a maximum temperature of 295 K. It has not yet been possible to perform other experiments, such as salt titrations and pH titrations, for this protein.

The mutant proteins gVp Tyr26-Cys and Phe73-Leu behave very similarly in solution and have a sample quality that is comparable to wt gVp. Samples of gVp Phe73-Leu isolated both from the soluble fraction and the insoluble fraction were also compared. The ^{1}H-NMR spectra of both samples were almost identical with a small difference in aggregation state in favour of the sample which was isolated from the soluble fraction.

The proteins gVp Tyr34-His, gVp Tyr41-Phe and gVp Tyr41-His clearly exhibit different properties from wt gVp. The ^{1}H-NMR spectra of these samples clearly indicate that the aggregation is much less severe than in wt gVp. The solution properties of these samples were therefore further studied by means of salt titrations, pH titrations and concentration dependency tests. Good quality ^{1}H-NMR spectra of gVp Tyr34-His were obtained up to a protein concentration of 3 mM. However, samples of gVp Tyr34-His were not as stable as wt gVp which made it more difficult to test its solution properties further. The p*K* of the histidine introduced at position 34 is lower than 5.0. An exact determination could not be made because of rapid protein denaturation below pH 5.

At low ionic strength, the gVp mutants Tyr41-His and Tyr41-Phe gave rise to ^{1}H-NMR spectra with relatively sharp resonances up to a concentration of approximately 5 mM (pH 5.1). At these concentrations, the proteins are stable for at least a few days at a

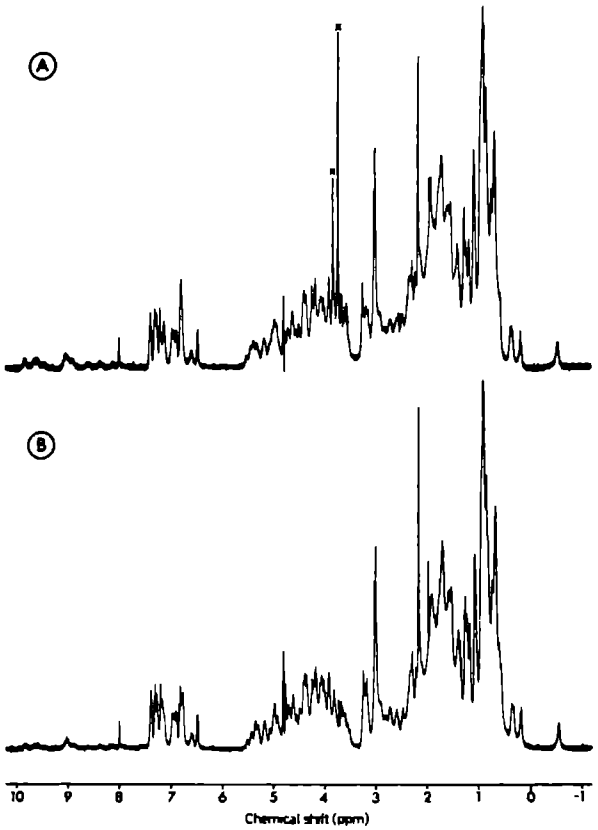


Figure 4.4 ^{1}H-NMR spectra of M13 wt gVp.
0.5 mM wt gVp pH 5.2 which was (A) subjected to a denaturation step with guanidine hydrochloride (the impurities introduced are indicated by an x) and (B) taken from the same batch as a control.

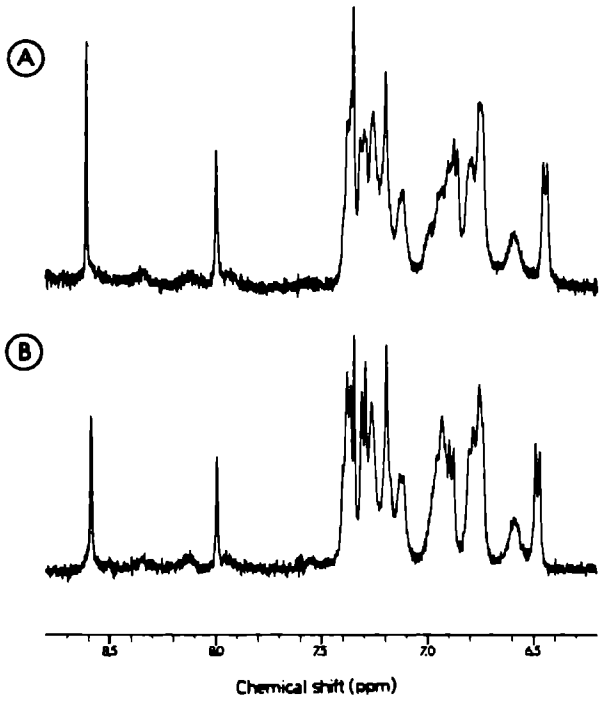


Figure 4.5 Concentration dependency of the aromatic region of the ^{1}H-NMR spectrum of M13 Tyr41-His gVp (pH 5.2).
M13 Tyr41-His gVp concentrations: (A) 4 mM (B) 0.5 mM.

temperature of 298 K. Increase of the temperature by only a few degrees results in rapid denaturation. Both samples are less sensitive to changes in salt and protein concentration than samples of wt gVp. In Figure 4.5., the aromatic spectrum of the sample Tyr41-His is shown at two different protein concentrations. The spectral alterations which can be observed are not as prominent as in the case of wt gVp and mainly involve a line broadening effect going to a higher concentration. The line widths of the individual resonances still remain relatively small at a concentration of 4 mM. The peaks belonging to the spin-system TyrIV in the specta of wt gVp are missing. The histidine introduced at position 41 in the Tyr41-His mutant has a p*K* of 6.5, indicative of a residue which is exposed to the solvent.

Sequence-specific assignment of the aromatic residues

Sequence specific assignments for the aromatic region of wt gVp were obtained by making a comparison of the TOCSY spectra of all aromatic mutant samples. The applicability of this approach is limited to the case where the spectra of wild-type and mutant protein are very similar. In Figure 4.6., the aromatic regions of the TOCSY spectra of wt gVp and of six mutant proteins are displayed. It can be seen that, apart from the absence of a few resonances, most of the aromatic spectra of mutant gVp are almost identical to that of wt gVp. Some resonances undergo small shifts but could nevertheless be easily identified.

In Table 4.1, a summary of the assignments obtained for wt gVp is presented. The distinction between δ and ε resonances, which has in part already been made previously (Garssen *et al.*, 1980; Alma *et al.*, 1981a; de Jong *et al.*, 1987b), was re-evaluated by means of a NOESY (D_2O) spectrum of wt gVp which showed NOE crosspeaks between the Hβ proton resonances and the Hδ ring proton resonances (data not shown). In the spectrum of the gVp mutant Tyr26-Cys, the crosspeaks connecting the resonances at 6.48 ppm and 6.88 ppm, which earlier had been designated as TyrI in wt gVp, are clearly missing (Figure 4.6B). The rest of the aromatic spectrum is hardly different from that of wt gVp and therefore TyrI can be ascribed to Tyr26. Similarly, in the spectrum of gVp Tyr34-His, the TyrV spin-system is missing (Figure 4.6C). The spectrum displays a downfield-shifted TyrIV, indicative of a decreased level of aggregation. This observation is in agreement with the improved spectral quality of gVp Tyr34-His in comparison to wt gVp.

The tendency to aggregate is even less in the gVp mutants Tyr41-Phe and Tyr41-His, in which the TyrIV

spin-system is missing. Now it can be concluded unambiguously that the residue which accounts most for the aggregation effect in wt gVp is the tyrosine residue at position 41 (Figures 4.6D and 4.6G). The aromatic spectrum of the mutant gVp Tyr61-His clearly lacks the broad resonances at 6.58 ppm and 6.75 ppm in wt gVp (Figure 4.6E). Tyr61 makes up spin-system TyrII and the only remaining tyrosine residue at position 56 can now by assigned to spin-system TyrIII by elimination.

Due to the availability of only one phenylalanine mutant, a complete assignment of all phenylalanine residues using the site-specific mutants could not be made. The spectrum of the gVp mutant Phe73-Leu can be superimposed on the spectrum of wt gVp except for the missing spin-system PheI (Figure 4.6F). The assignment of the phenylalanine residues could however be completed using the sequence-specific resonance assignment procedure, which could be applied to the gVp mutant Tyr41-His (Wüthrich, 1986; van Duynhoven *et al.*, 1990). The reported sequential assignment of a β-loop of this molecule, encompassing residues 13-31, included the aromatic residues Phe13 and Tyr26. Because of the similarities between wt gVp and mutant gVp Tyr41-His, the sequential assignments of gVp Tyr41-His were used to complete the assignment of the aromatic resonances in wt gVp. Indeed, the assignment of Tyr26 in gVp Tyr41-His agrees with the assignment based on mutant studies. Phe13 had not been assigned before and can now be attributed to the PheIII spin-system. Consequently, spin-system PheII can be assigned to Phe68 by elimination.

Table 4.1 Resonance assignments for the aromatic region of M13 wt gVp at very low ionic strength, pH 7.1, 298 K.
Chemical shifts are expressed relative to 2,2-dimethyl-2-silapentane-5-sulfonate. The residue-specific assignments according to Alma *et al.* (1981a) appear in parentheses.

Residue		chemical shift		
		δ	ε	ξ
Tyr26	(TyrI)	6.88	6.48	-
Tyr34	(TyrV)	6.96	6.75	-
Tyr41	(TyrIV)	7.16	6.82	-
Tyr56	(TyrIII)	7.12	6.78	-
Tyr61	(TyrII)	6.58	6.75	-
Phe13	(PheIII)	6.92	7.25	7.25
Phe68	(PheII)	7.39	7.17	6.95
Phe73	(PheI)	7.31	7.39	7.31
		δ2	ε1	
His64		7.19	7.97	

Binding of wild-type gene V protein and aromatic mutant gene V proteins to spin-labelled oligonucleotides

Binding of wt gVp to oligonucleotides causes upfield shifts of specific aromatic resonances. Significant shifts have been observed for one phenylalanine and two tyrosine residues (de Jong *et al.*, 1987b; King and Coleman, 1987). However, for one of the tyrosine residues the shift does not only depend on the degree of saturation of the protein with ssDNA but also on the total protein concentration (de Jong *et al.*, 1987b). Therefore, it has been concluded that only two aromatic residues are involved in binding to ssDNA (de Jong *et al.*, 1987b; King and Coleman, 1987). These can now definitely be identified as Tyr26 and Phe73.

Apart from binding normal oligonucleotides to gVp, it is also possible to probe the ssDNA binding domain and its surroundings with spin-labelled oligonucleotides (de Jong *et al.*, 1989b). The spin-labelled ligand bound to a macromolecule selectively broadens the resonances which are in close vicinity of the spin-label. Difference spectra computed from spectra obtained with and without spin-labelled ligand will thus display the resonances originating from the residues which are situated in the region of the binding domain.

The changes observed in the aromatic part of the spectrum upon binding of spin-labelled oligonucleotides to M13 wt gVp are congruent with the changes observed in binding of unmodified oligonucleotides. This is shown in Figure 4.7, in which a comparison is made of the aromatic spectra of M13 wt gVp in the absence and presence of small amounts of *$d(A)_3$*. The aromatic resonances of Phe73 and Tyr26 are erased from the spectrum. The difference spectrum also reveals some shifts that are caused by the change in sample conditions upon addition of the spin-labelled DNA. Very prominent in this respect is the shift of spin-system TyrIV which is now assigned to Tyr41. This effect can be reduced by diluting the sample which further indicates that Tyr41 does not form a part of the DNA binding domain.

Similar experiments were performed on mutant gVp samples. A few examples are shown in Figure 4.7. In gVp Tyr34-His, Tyr26 and Phe73 show up in the difference spectra. Residue Tyr41 shifts as well but to a lesser extent than for the wt gVp. The binding domain seems also to be unaffected in the gVp mutants Tyr41-His and Tyr41-Phe. Despite the absence of the Tyr41 residue and the reduced aggregation, it can be seen that the shift effects did not vanish: apart from small shifts of the residue (histidine/phenylalanine) introduced at position 41, other shifts could be observed, such as those of Tyr34 and Phe13. These residues were identified in two-dimensional difference spectra (data not shown). The mutants Phe73-Leu and Tyr26-Cys still bind to the spin-labelled oligonucleotide despite the absence of one of the two important aromatics. In the difference spectra of mutant gVp Phe73-Leu the tyrosine residue at position 26 can be seen. The difference spectra of the gVp Tyr26-Cys mutant show the Phe73 residue. These data suggest that the major part of the DNA binding domains of these mutants is still intact.

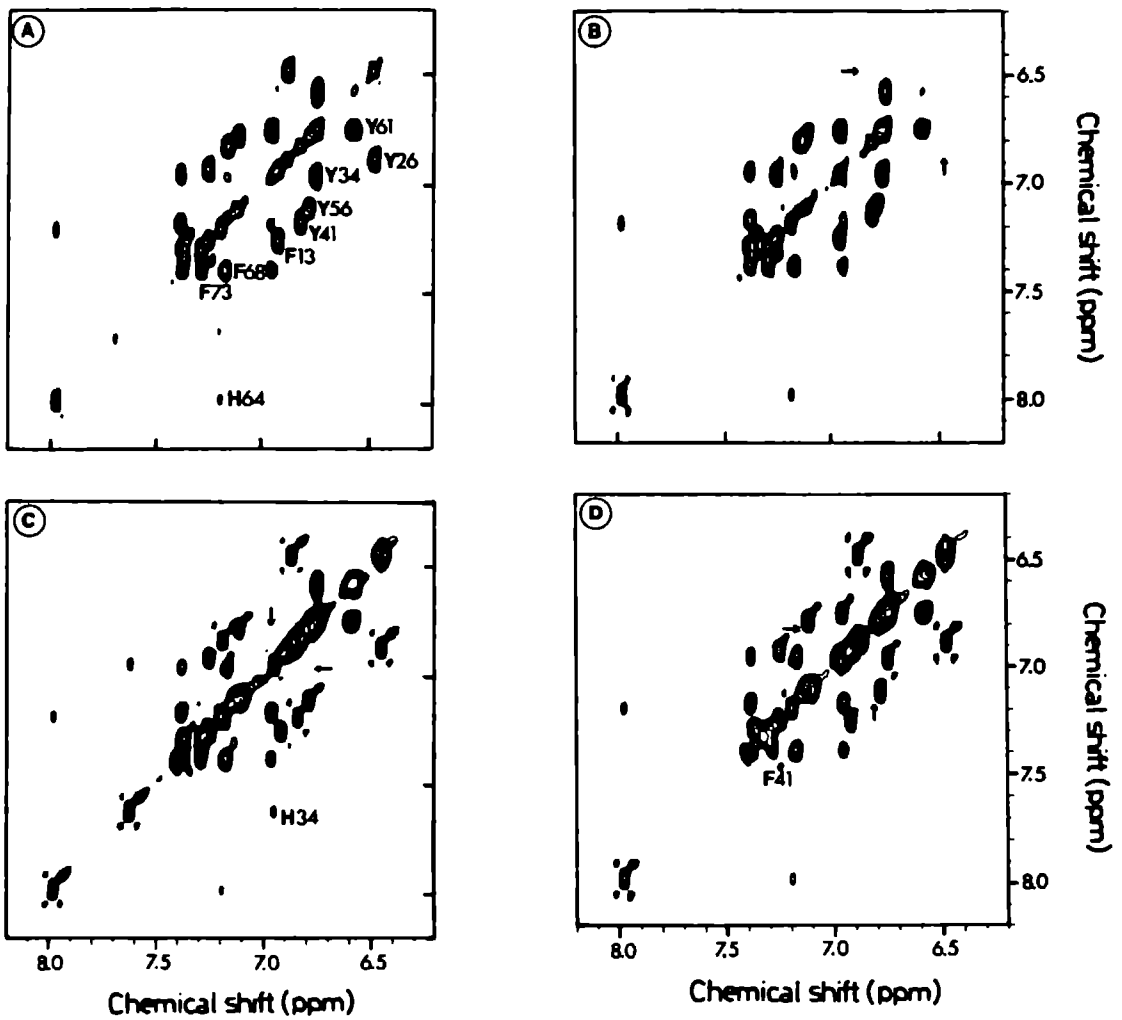


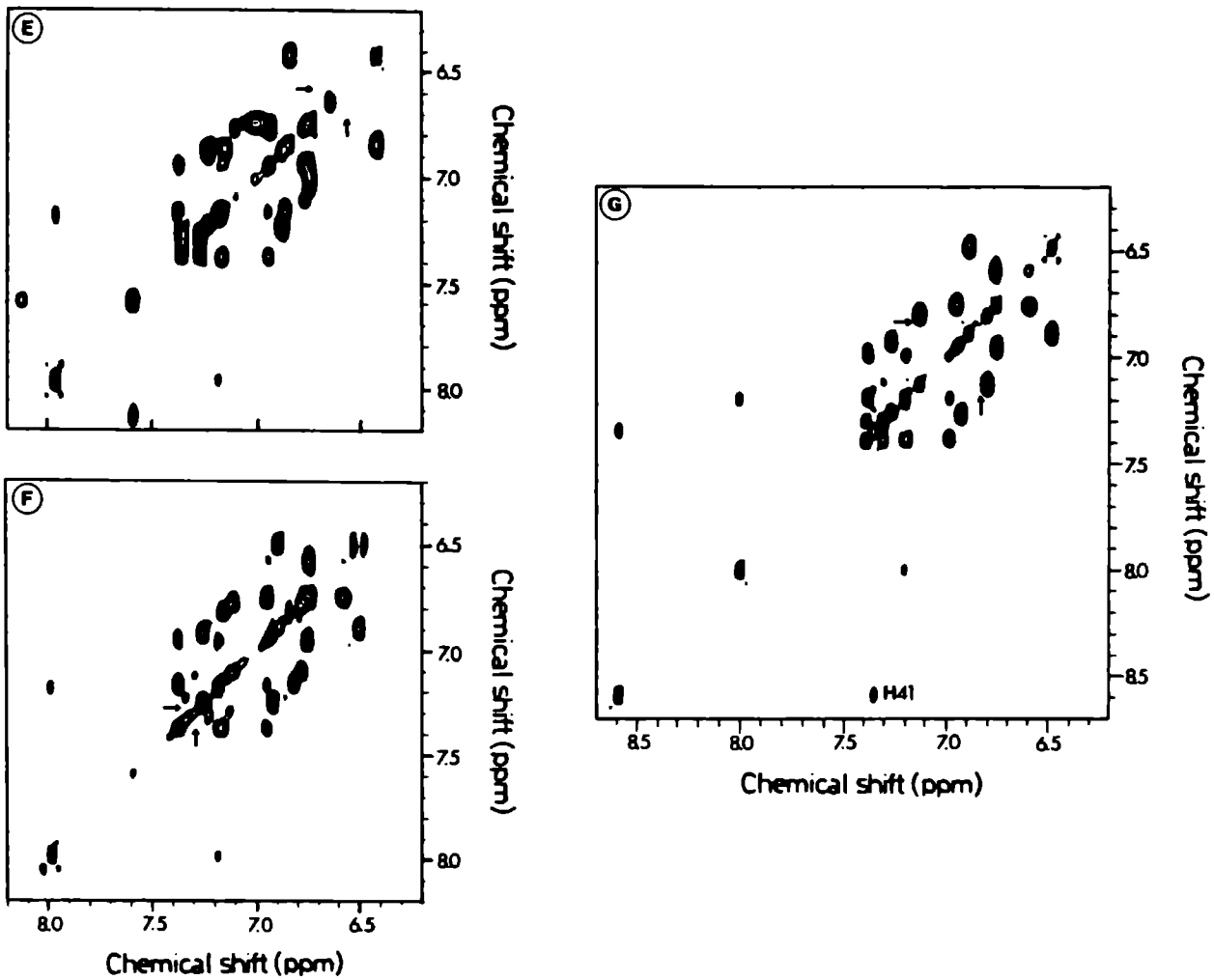
Figure 4.6 Aromatic regions of TOCSY spectra of M13 wt gVp and a variety of aromatic gVp mutants. All spectra were recorded at 298 K unless otherwise stated. The mixing times used were approximately 30 ms. Sequence-specific assignments are indicated in the figure near (one of) the crosspeaks connecting the resonances of the spin-system see Figure 4.3).
A) 0.5 mM wt gVp pH 7.1; (B) 1.0 mM gVp Tyr26-Cys pH 7.0; (C) 3 mM gVp Tyr34-His pH 5.2; (D) 4 mM gVp Tyr41-Phe pH 5.2; E) 0.5 mM gVp Tyr61-His at 295 K, pH 7.0; (F) 0.5 mM gVp Phe73-Leu pH 5.2; (G) 1 mM gVp Tyr41-His pH 5.1.

DISCUSSION

Optimized conditions have been obtained for measurement of the ^{1}H-NMR spectrum of M13 wt gVp. These conditions should also be used for measurements of several mutant gVp samples we have at our disposition. From the results presented, it can be concluded that NMR studies on wild-type and mutant M13 gVp can best be conducted at very low ionic strength and at pH 5.1.

M13 wt gVp has a high tendency to aggregate. The aggregation effect could be evaluated by considering the aromatic region of the ^{1}H-NMR spectrum of wt gVp under different solution conditions. Upon increase of the protein and/or salt concentration, two effects occur: the first is a broadening of all spectral lines; the second is an upfield shift of the peaks belonging to a single tyrosine residue, which has now been assigned to Tyr41. In this respect, the behaviour of M13 wt gVp is very similar to that of the previously studied IKe gVp. IKe gVp also contains a tyrosine spin-system which is very sensitive to changes in solution conditions (de Jong *et al.*,1987b). This particular tyrosine spin-system in IKe gVp has been assigned to Tyr42, by means of sequential assignment procedures (de Jong *et al.*, 1989a). Because of the similar sequences of both proteins and the similarity in behaviour of this specific residue, it had already been suggested that this residue in M13 wt gVp is Tyr41. Using missense mutants, we have been able to show that this is indeed the case.

The analysis of the spin-patterns in the aromatic region of the ^{1}H-NMR spectrum of M13 wt gVp by means of TOCSY spectra is in agreement with the



analysis made by Alma *et al.* (1981a). However, it differs slightly from the data presented by King and Coleman (1987). Their NMR samples contained between 1-1.8 mM wt gVp in the presence of 25 mM phosphate buffer pH 7.6. NOESY experiments were conducted at temperatures ranging between 298 K and 303 K. The major difference between their data and ours is the TyrIV spin-pattern: in their spectra it exhibits four different chemical shift positions for the protons of the ring, indicating that the residue is immobilized; in our case, it appears to be rapidly flipping around the Cβ-Cγ bond. This indicates that at their conditions, aggregation of the protein has occurred which led to the immobilisation of the tyrosine ring.

The solution properties of the gVp mutants we have studied differ significantly from one another. One factor affecting the quality of the protein is the protein isolation procedure used. For a few mutant clones, the expressed proteins accumulate in so-called 'inclusion bodies' in the cells and were isolated via a denaturation step. It could be shown that, although the proteins isolated via this procedure did have a correct fold, the level of aggregation was clearly increased relative to the proteins which were isolated from the soluble fraction of the cell extract.

However, the solubility of the various mutants cannot be classified according to the isolation procedure used. GVp mutant Tyr61-His, which was isolated from 'inclusion bodies' of the *Escherichia coli* cells, is hardly soluble at concentrations at which NMR spectra can be recorded. The spectrum of the aromatic region contains an extra crosspeak in comparison to wt gVp. A possible explanation is that the Tyr41 is slowly flipping, indicative of the severe aggregation and low solubility. The increased aggregation of this particular mutant in comparison to wt gVp cannot be explained as yet.

The other set of mutants can be divided into two categories. The first category comprises mutants such as gVp Phe73-Leu and gVp Tyr26-Cys which have properties similar to wt gVp. The second category consists of the mutants which have improved solubility characteristics, such as gVp Tyr34-His, gVp Tyr41-His and gVp Tyr41-Phe. Increase of protein concentration and/or salt concentrations in wt gVp and mutants gVp Tyr26-Cys and gVp Phe73-Leu results in broadening of all spectral lines combined with the prominent shift of the Tyr41 resonances. The specific shift of the Tyr41 residue can also be seen upon addition of normal

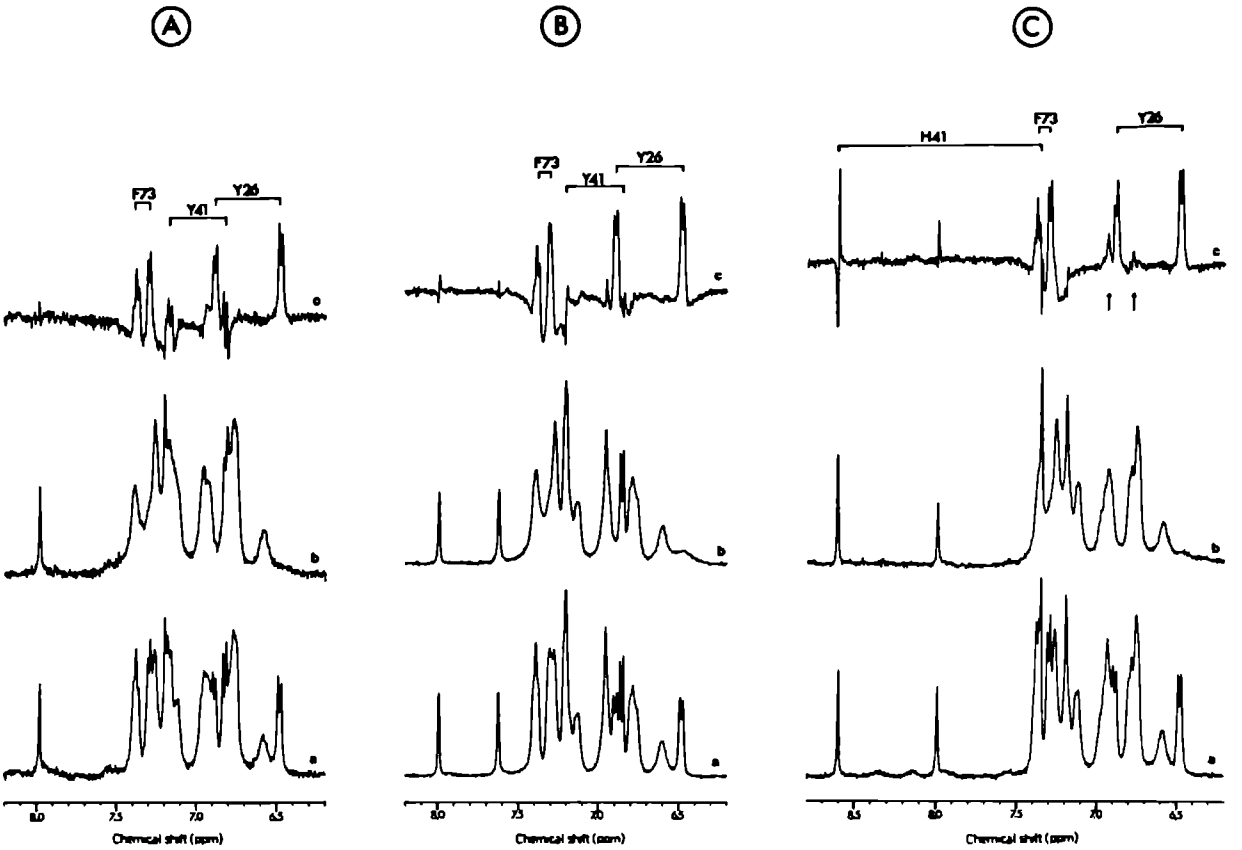


Figure 4.7 Examples of spectra of the aromatic region of wt gVp (A), gVp Tyr34-His (B) and gVp Tyr41-His (C), recorded in the absence and presence of the spin-labelled oligonucleotide *d(A)$_3$*. The small arrows in the gVp Tyr41-His spectrum (C) indicate the presence of small shift effects of Phe13 (left arrow) and Tyr34 (right arrow). (a) Spectrum in the absence of spin-labelled oligonucleotide; (b) spectrum after addition of 1/25 equivalent *d(A)$_3$* (1), of 1/45 equivalent *d(A)$_3$* (2) and 1/12 equivalent *d(A)$_3$* (3), respectively; (c) the difference between (a) and (b).

oligonucleotides as well as spin-labelled oligonucleotides to both wild-type and these mutant proteins. It is very likely that the Tyr41 residue plays a crucial role in the specific recognition of gVp dimers. On the basis of model study, Brayer and McPherson have proposed that hydrophobic interactions between the protein surfaces of two dimers may form the physical basis for the observed co-operativity in the binding to ssDNA. The Tyr41 residue is part of one of the suggested interaction sites.

From these considerations, it can be suggested that the specific aggregation has to be explained on the basis of a set of interactions between dimers responsible for the co-operativity in the binding to ssDNA. Substitution of the crucial Tyr41 residue resulted in changes in the solubility characteristics of the protein. The strong interaction between the gVp dimers is clearly diminished in the gVp mutants Tyr41-His and Tyr41-Phe. This result further supports the hypothesis that the Tyr41 residue plays a dominant role in the interdimer interactions. From the DNA binding experiments and the TOCSY spectra of wt gVp at different conditions in combination with two-dimensional NMR experiments (data not shown), it can be seen that, for example, the tyrosine residue at position 34 undergoes a minor shift under a change in solution conditions. This implies that this residue also might be part of the possible dimer-dimer interaction surface. However, it can also be explained as an intradimer effect in which the Tyr41 residue affects the Tyr34 residue.

The solubility and the spectral appearance of the gVp mutant Tyr34-His has improved in comparison to the wt gVp. In line with the considerations made above this can be either an intradimeric or an interdimeric effect. The reduced stability of the Tyr34-His mutant cannot be explained as yet.

The differences in solution properties could be explained in terms of a specific residue by virtue of the complete sequence-specific assignment of all aromatic residues. Previously, sequence-specific assignments of wt gVp have been obtained for all aromatic residues on the basis of NOESY spectra and the crystallographic data of wt gVp (King and Coleman, 1987). Our results differ from those reported earlier with respect to the assignments of Tyr34 and Tyr41. It is noted that the

assignment of Tyr41 had been corrected before on the basis of a mutant protein sample of which, to our knowledge, details never have been published (King and Coleman, 1988). This indeed indicates the necessity for unambiguous assignments which is difficult when only founded on a crystallographically determined structure. NMR studies on gVp Tyr41-His have also revealed that the secondary structure of the sequence-specifically assigned part of the molecule differ from the structure postulated for the same region on the basis of the X-ray crystallographic studies (van Duynhoven *et al.*, 1990). This again emphasizes the potential danger of making assignments which are solely based on crystallographic data.

Although the availability of aromatic mutant samples was limited to only 5 out of a total of 8 unassigned aromatic residues, a complete assignment has been made. For mutant gVp Tyr41-His of which the solubility characteristics are significantly better than those of wt gVp, it became possible to perform two-dimensional NMR measurements in H_2O solution, a prerequisite for application of the sequence-specific resonance assignment approach (van Duynhoven *et al.*, 1990). These measurements were very hard to perform on wt gVp because the aggregation was still too severe to obtain spectra with a sufficiently suppressed H_2O resonance and an interpretable spectral appearance. The sequence-specific assignments made of gVp Tyr41-His included Phe13 which was enough to complete the assignment of the aromatic spectrum.

The results of binding experiments of the spin-labelled oligonucleotides to wt gVp and to aromatic mutant proteins are in line with results of NMR studies reported previously on M13 wt gVp and the gVp of the evolutionarily related phage IKe (de Jong *et al.*, 1987b; King and Coleman, 1987). From the binding experiments, it can be concluded that only two aromatic residues, namely Tyr26 and Phe73, are involved in the binding to oligonucleotides. These data are not in agreement with model building proposals of Brayer and McPherson (1984a) which were based on their reported crystal structure. In their proposals, it is suggested that the aromatic residues Tyr34 and Tyr41 are also involved in binding to ssDNA. Our data and those previously published by King and Coleman (1987) clearly suggest that these residues are not involved in the binding to oligonucleotides. NMR binding studies with longer ssDNA fragments which may be more relevant with respect to complex formation under co-operative binding conditions have not altered that view (King and Coleman, 1988). The Tyr41 residue instead plays an important role in interdimer interactions which may be relevant to co-operativity in binding of gVp to ssDNA which we have discussed above.

Recently, a new model has been proposed for the complex of gVp with ssDNA (Hutchinson *et al.*, 1990). In the model, Tyr26, Phe73 and Tyr34 interact with nucleotide bases. It had been suggested that Tyr34 interacts with nucleic acid in the n=4 binding mode but not in the n=3 binding mode. Thus far, our studies with spin-labelled oligonucleotides indicate that, in the case of oligonucleotide binding, Tyr34 is neither part of nor close to the DNA binding domain.

CHAPTER 5

NUCLEOTIDE SEQUENCE OF THE GENOME OF THE FILAMENTOUS BACTERIOPHAGE I2-2

MODULE EVOLUTION OF THE FILAMENTOUS PHAGE GENOME

A.P.M. STASSEN, E.F.P.M. SCHOENMAKERS, M. YU, J.G.G. SCHOENMAKERS, and R.N.H. KONINGS (1992) *J. Mol. Evol.* **34**: 141-152.

ABSTRACT

The nucleotide sequence of the circular single-stranded genome of the filamentous *Escherichia coli* phage I2-2 has been determined and compared with those of the filamentous *E. coli* phages Ff (M13, f1 or fd) and IKe. The I2-2 DNA sequence comprises 6744 nucleotides; 139 nucleotides less than that of the N- and I2-plasmid specific phage IKe, and 337 (336) nucleotides more than that of the F-plasmid specific phage Ff. Nucleotide sequence comparisons have indicated that I2-2, IKe and Ff have a similar genetic organization, and that the genomes of I2-2 and IKe are evolutionarily more closely related than those of I2-2 and Ff. The studies have further demonstrated that the I2-2 genome is a composite replicon, composed of only two thirds of the ancestral genome of IKe. Only a contiguous I2-2 DNA sequence of 4615 nucleotides encompassing not only the coat protein and phage assembly genes, but also the signal required for efficient phage morhogenesis, was found to be significantly homologous to sequences in the genomes of IKe and Ff. No homology was observed between the consecutive DNA sequence that contains the origins for viral and complementary strand replication and the replication genes. Although other explanations cannot be ruled out, our data strongly suggest that the ancestor filamentous phage genome of phages I2-2 and IKe has exchanged its replication module during evolution with that of another replicon, *e.g.* a plasmid that also replicates via the so-called 'rolling circle' mechanism.

INTRODUCTION

A major characteristic of filamentous bacteriophages is that infection of their host cell occurs via attachment of a specific phage end, *i.e.* the one at which the coat proteins encoded by genes III and VI are located, to the tip of a conjugative pilus present at the surface of the host cell (Caro and Schnös, 1966). This pilus is composed of filamentous polymers of identical protein subunits (pilin), generally encoded by conjugative plasmids (Marvin and Folkhard, 1986). On the basis of their plasmid (pilus) specificity, filamentous phages have been subdivided into several groups which differ both in phage morphology and protein composition. The best studied filamentous phages are the IncF plasmid specific phage Ff (M13, f1 and fd) (for a review, see Model and Russel, 1988) and the IncI2- and IncN-plasmid specific phage IKe (Khatoon *et al.*, 1972; Bradley *et al.*, 1983; Peeters *et al.*, 1985). They both have *Escherichia coli* as a host. Filamentous phages with other pilus- and host specificity are Pf3, a *Pseudomonas* phage, and C-2, a *Salmonella* phage. They are specific for IncP-1- and IncC-plasmids respectively (Stanisich, 1974; Bradley *et al.*, 1982).

The filamentous *E. coli* phage I2-2, which has been isolated from Pretoria sewage, is specific for pili encoded by conjugative plasmids of the IncI2-, IncN-, and IncP-incompatibility groups (Coetzee *et al.*, 1982; Bradley *et al.*, 1983). Its serological relationship with phage IKeh, a host range mutant of IKe that also infects *E. coli* cells containing plasmids of the IncP incompatibility group, is indicative of an evolutionary relationship between these two phages (Grant *et al.*, 1978; Bradley *et al.*, 1983). Based upon this relationship it has been suggested that I2-2 is a naturally occuring host range mutant of IKe (Bradley *et al.*, 1983).

Because we are interested in the evolutionary relationships of filamentous phages and because

comparison of sequences of evolutionary related genes or genomes might provide important clues about the significance of particular nucleotide and/or amino acid sequences with respect to their function, the nucleotide sequence of the circular single-stranded genome of I2-2 has been determined and compared with those of Ff and IKe. From these comparisons it became apparent that I2-2, IKe and Ff have a similar genetic organization and that the genomes of I2-2 and IKe are evolutionarily more closely related than those of I2-2 and Ff. The homology between the nucleotide sequences of the genomes of I2-2 and IKe or Ff is, however, only partial and restricted to the *cis*-acting nucleotide sequence required for phage assembly and the genes coding for the coat- and phage assembly proteins.

MATERIALS AND METHODS

Bacteria, bacteriophages and plasmids

Bacteriophage I2-2 and its host *E. coli* JE2571[N3] (*leu, thr, str, fla, pil,* [N3, *tet, strep, sulf*]), were kindly provided by Dr. D.E. Bradley, St. John's, Newfoundland (Watanabe *et al.*, 1964; Bradley *et al.*, 1982). For propagation of recombinant pKUN19 phagemids (Konings *et al.*, 1987), *E. coli* JM83 (*ara,* Δ(*lac-proAB*), *rpsL, φ80, lacZ*ΔM15) (Yanisch-Perron *et al.*, 1985) was used. For the production of single-stranded phagemid DNA by recombinant pKUN19 phagemid harbouring cells (*E. coli* JM101 (*supE, thi,* Δ(*lac-proAB*), [F', *traD36, proAB, lacI*^{q}*Z*ΔM15])), the helper phages IR1 (Enea and Zinder, 1982), Mike (Konings *et al.*, 1986) or MikeΔ (Konings *et al.*, 1987) were used. The single-stranded bacteriophage vectors M13mp18 and M13mp19, which were propagated on *E. coli* JM101, have been described previously (Yanisch-Perron, 1985).

Table 5.1 Synthetic oligonucleotides and *Hae*III fragments of I2-2 RF that have been used as nucleotide sequence analysis primers for the viral strand of phage I2-2.

Primer	Nucleotide Positions in the I2-2 genome
5'- dCCCAGCCTAATTTACGGGC -3'	226 - 244
5'- dGTGAACTTTCGGTAATTG -3'	394 - 411
5'- dGCGCACAGGAGAATTG -3'	1070 - 1085
5'- dACTGTTATTGCCTGGGTA -3'	1485 - 1502
5'- dTTAACGCTTGCAGCC -3'	1895 - 1909
5'- dGGCAATGATAGCGTG -3'	3020 - 3033
5'- dGACGATACGATTCTAG -3'	3406 - 3421
5'- dGGCACAGACAGTTCC -3'	5515 - 5529
5'- dACATCTTCCTTTGTC -3'	5757 - 5771
5'- dCGCTATCTGGTAGTC -3'	6558 - 6572
*Hae*III fragments	4490 - 4708
	4709 - 4832
	4833 - 5229
	5230 - 91

Enzymes and chemicals

Restriction endonucleases were purchased from Boehringer-Mannheim (GER), New England Biolabs (Beverly, USA), and Bethesda Research Laboratories (Gaithersburg, USA). T4 DNA ligase, *E. coli* DNA polymerase I (Klenow fragment), and calf intestine alkaline phosphatase were purchased from Boehringer-Mannheim (GER). All enzymes were used according to the manufacturers instructions.

The 2',3'-dideoxyribonucleosidetriphosphates and 2'-deoxyribonucleosidetriphosphates were purchased from Pharmacia (Uppsala, Sweden) and Boehringer-Mannheim (GER), respectively. α-^{32}P-labelled deoxyadenosine triphosphate and α-^{35}S-labelled thio-deoxyadenosine triphosphate were obtained from Amersham (Buckinghamshire, UK).

The oligonucleotides, used for nucleotide sequence analysis of the viral strand, are listed in Table 5.1. They were synthesized either by a Cyclone Plus DNA Synthesizer (MilliGen/ Biosearch) or by an Applied Biosystems DNA Synthesizer.

I2-2 replicative form DNA and single-stranded DNA

Replicative form DNA (RF) was isolated from I2-2 infected *E. coli* JE2571[N3] cells, using the alkaline lysis method of Birnboim and Doly (1979) as described previously by Konings *et al.* (1987). The procedures for the preparation of a high-titer I2-2 phage stock and the isolation of viral single-stranded DNA, have been described previously (Konings *et al.*, 1987).

Cloning of I2-2 DNA and nucleotide sequence analysis

Basic DNA manipulations, which will not be elaborated here, were mainly performed according to Sambrook *et al.* (1989).

A library of the I2-2 genome was constructed by random cloning of *Hae*III restriction fragments of I2-2 RF in the *Sma*I site of either pKUN19, M13mp18 or M13mp19. After digestion with *Sau*3AI, the *Hae*III restriction fragments were subcloned in the *Bam*HI restriction site of pKUN19.

The nucleotide sequence of the cloned *Hae*III and *Sau*3AI fragments was determined by the chain termination method developed by Sanger *et al.* (1977). For the analyses also the single-stranded I2-2 genome was used as a template. As sequencing primers synthetic oligonucleotides and *Hae*III fragments of I2-2 RF (Table 5.1) were used. To prevent band compression the dGTP analog 7-deaza-dGTP, instead of dGTP, was used.

The nucleotide sequence was analysed by computer programmes developed by Staden (1980, 1982) and Wilbur and Lipman (1983), and the computer programme package IntelliGenetics Suite (release 5.37) of IntelliGenetics, Inc. (California, USA). To search for similar nucleotide sequences, the EMBL (release 23) and Genbank (release 60) Nucleic Acid Databases were scanned. The sequence data are stored in the EMBL, Genbank and DDBJ Nucleotide Sequence Databases under the accession number X14336.

To estimate the evolutionary relationship between

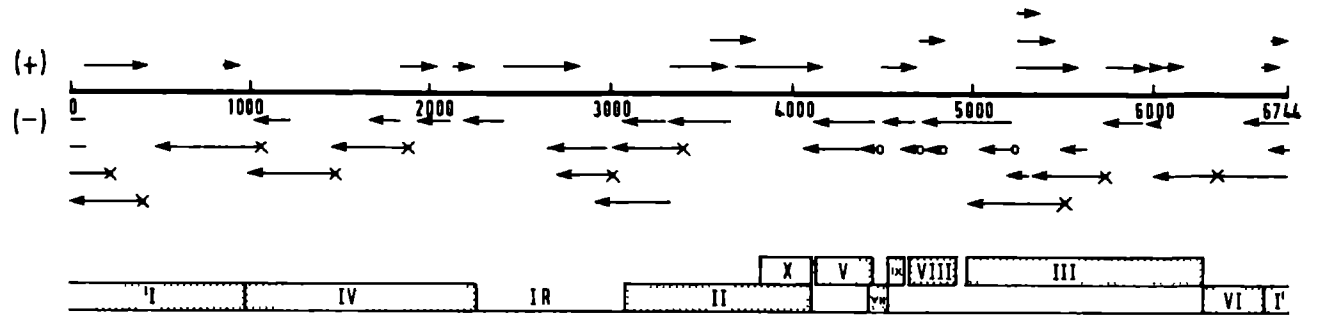


Figure 5.1 Positions of individual gel readings in the final nucleotide sequence of the I2-2 genome. The axis represents the nucleotide map of the I2-2 genome, starting at the unique *Hin*dIII recognition site, numbered in the 5'-3' direction of the viral strand. Arrows pointing to the right and left refer to nucleotide sequences established for the viral and complementary strand, respectively. The length of the sequence derived is represented by the length of the arrow. Nucleotide sequences indicated by ---> or <--- are obtained from cloned DNA, while those indicated by <--x and <--o are obtained from nucleotide analysis of viral DNA primed with synthetic oligonucleotides or *Hae*III restriction fragments, respectively. The linear genetic map of the I2-2 genome is given in the lower portion of the diagram. Genes are denoted by Roman numerals; IR refers to the major intergenic region in which the origins for viral and complementary strand synthesis and the signal required for efficient phage assembly are located.

he homologous genes (I, III, IV, VI, VII, VIII and IX) of 2-2, IKe (Peeters *et al.*, 1985) and Ff (van Wezenbeek *et al.*, 1980), the extent of synonymous (Ks = synonymous substitutions/synonymous sites) and non-synonymous amino acid substitutions (Ka = amino acid substitutions/amino acid sites) was calculated according to Miyata and Yasunaga (1980) and corrected for back mutations by the method of Jukes and Cantor 1969). Triplet insertions/deletions were excluded from he comparisons. The Ks and Kas calculated are thus overestimates.

RESULTS AND DISCUSSION

Nucleotide sequence of I2-2 DNA

The sequence analysis of the I2-2 genome was initiated by sequencing of I2-2 *Hae*III fragments andomly cloned into phagemid pKUN19. Because hree *Hae*III fragments were longer than 1000 base-pairs, and due to subcloning difficulties, several gaps emained. These gaps were filled in by using the viral genome as a template with synthetic oligonucleotides or purified *Hae*III fragments as sequencing primers (*vide supra*). Furthermore, the nucleotide sequence of several *Sau*3AI subclones of *Hae*III fragments was determined. The sequencing strategy followed is outlined in Figure 5.1.

The majority of the I2-2 sequence was established via sequence analysis of both the viral and complementary strand. On the average each nucleotide is supported by at least three separate gel readings. We herefore feel confident that the nucleotide sequence is presented in Figure 5.2 is correct. As nucleotide number 1 we have chosen the first nucleotide (A) of he unique *Hin*dIII restriction endonuclease cleavage site. The circular single-stranded genome of I2-2 is 5744 nucleotides long. This is 139 nucleotides less than that of IKe (Peeters *et al.*, 1985), and 337 (336) nucleotides more than that of Ff (Beck *et al.*, 1978; van Wezenbeek *et al.*, 1980; Beck and Zink, 1981; Hill and Petersen, 1982).

Gene identification

The five coat protein genes (*i.e.* genes III, VI, VII, VIII and IX), as well as the two phage assembly genes (I and IV) and the phage morphogenetic signal located immediately distal to gene IV, have been identified by alligning of the I2-2 sequence with that of phage IKe. The respective gene products are not only highly homologous but also of similar lengths (Figure 5.2). Although no significant homology was observed between the open reading frames II, X, and V of I2-2 and the DNA replication genes II, X, and V of IKe, as demonstrated here, we feel confident to conclude that the proteins encoded by these open reading frames fulfil an identical function in the DNA replication process. This conclusion is partly based on the observation that the biochemical and physical properties (DNA binding properties, amino acid composition and molecular weight) of the I2-2 encoded single-stranded DNA binding protein are in accordance with those of the protein encoded by open reading frame V (data not shown). Furthermore genetic and biochemical studies have unambiguously demonstrated that the filamentous phage genome codes for three different functions that are indispensable for the phage (DNA) replication process. One of these functions is encoded by gene II and indispensable for initiation and termination of viral strand replication. The second function is encoded by gene X, a gene that completely overlaps with the 3' terminal region of gene II, and is somehow involved in the regulation of the synthesis of the complementary strand. The third function is the single-stranded DNA binding protein encoded by gene V that, at a late stage of the infection cycle, is indispen-

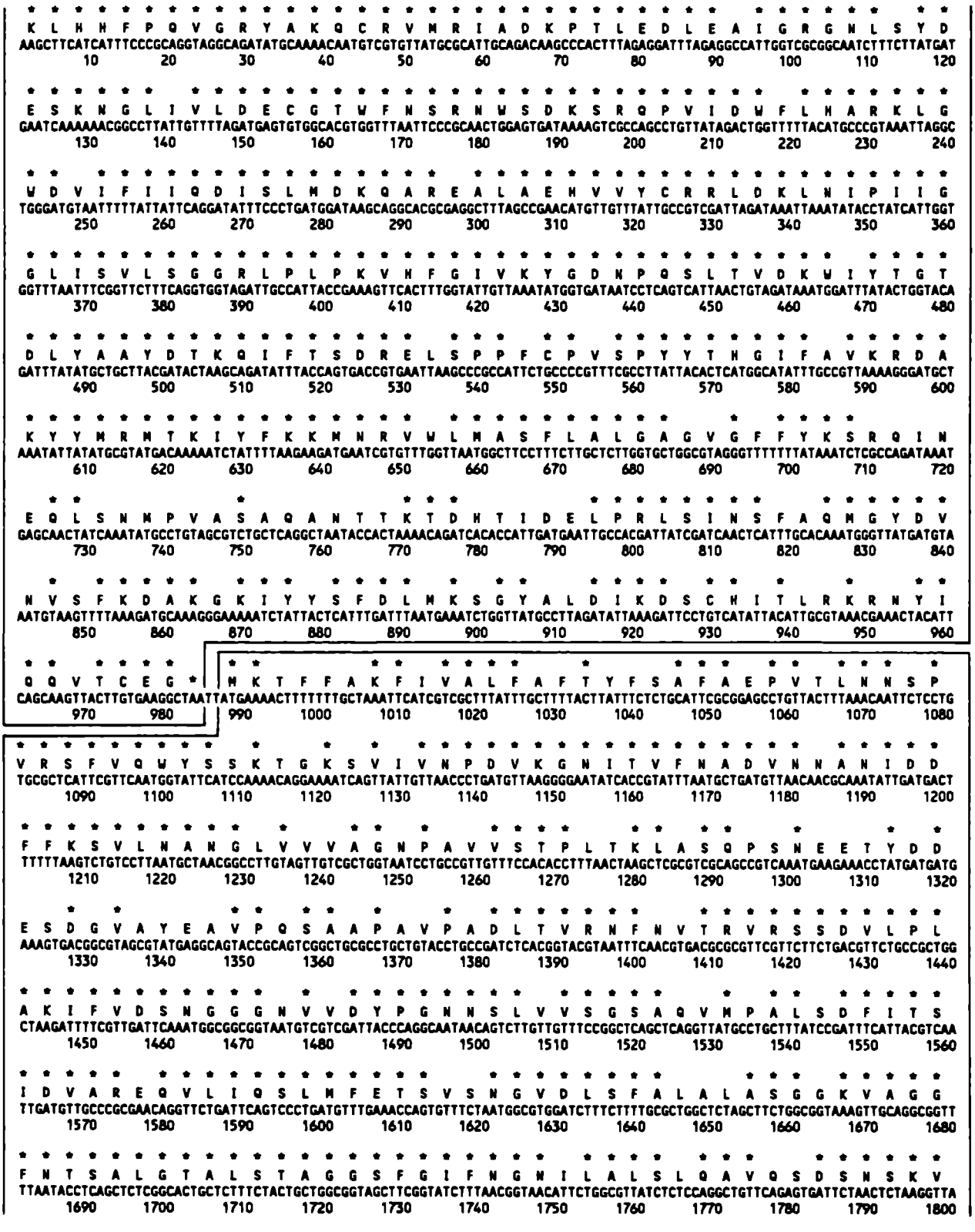


Figure 5.2 Nucleotide sequence of the single-stranded genome of the N/I2/P-specific filamentous *E. coli* phage I2-2.

Numbering is in the 5'-3' direction of the viral strand and begins at the first nucleotide of the unique *Hin*dIII recognition site. The genes I to X are boxed and the amino acid sequences of their encoded proteins are shown. For clarity gene VII is boxed with a broken line. The amino acid residues that have an identical position in the corresponding gene products of I2-2 and IKe are indicated by asterisks. Termination codons are denoted by asterisks above the nucleotide sequence. IR refers to the major intergenic region in which the origins for viral and complementary strand synthesis and the signal required for efficient phage assembly are located.

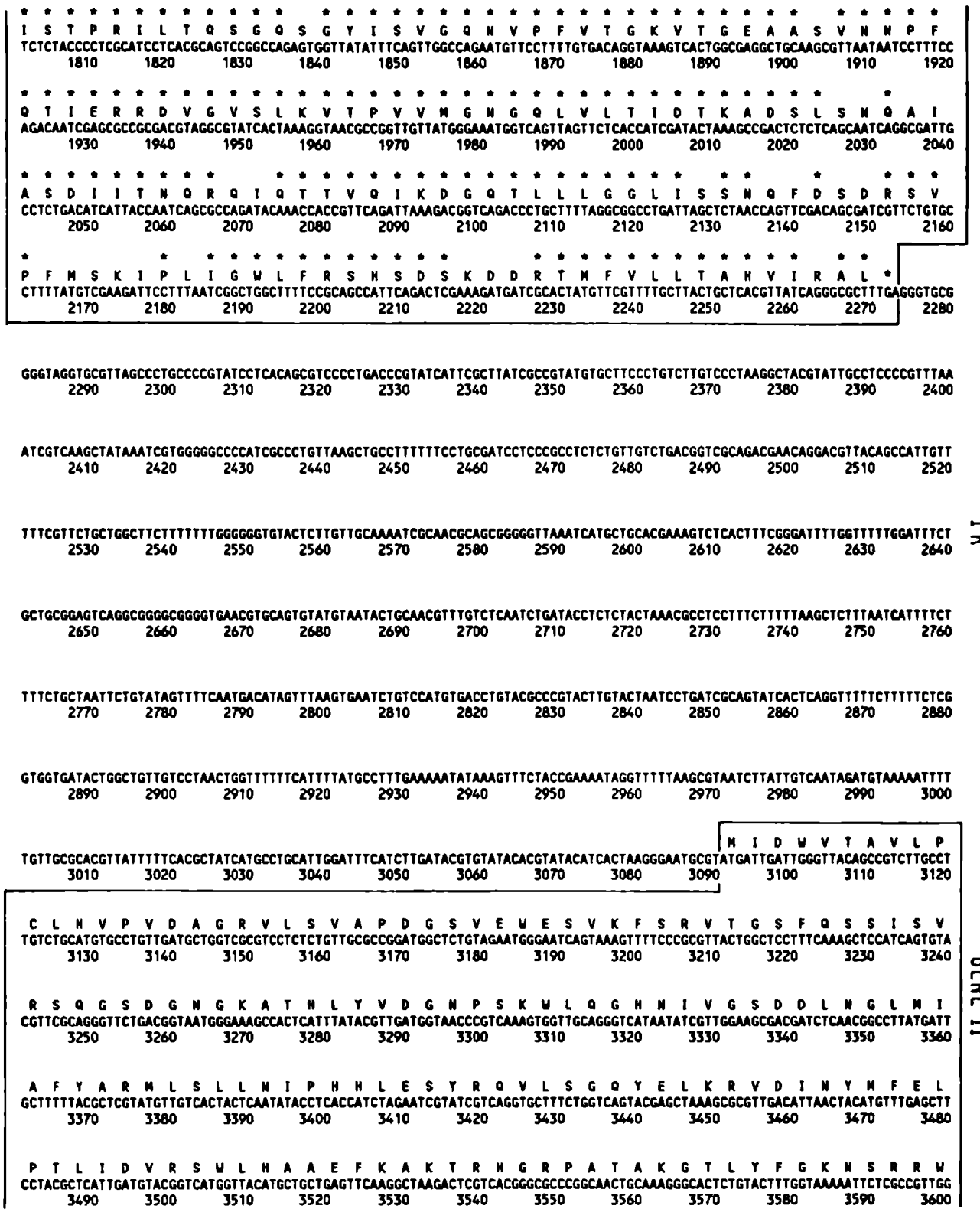


Figure 5.2 Continued

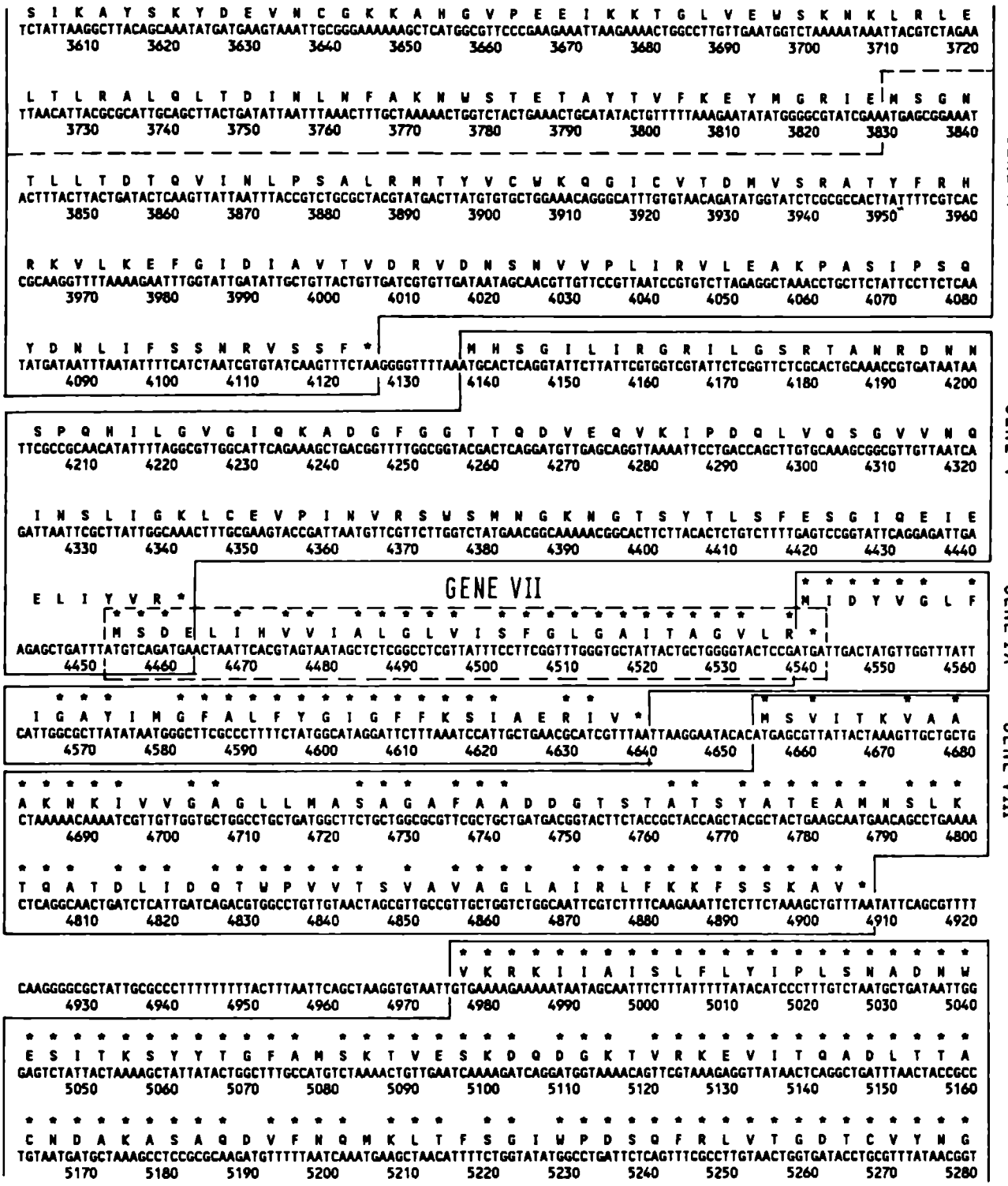


Figure 5.2 Continued

sable for the sequestering of the newly synthesised viral strand from the DNA replication cycle. Because the only DNA region that is left in the I2-2 genome that might code for these proteins is located between the major intergenic region and gene VII (Figure 5.3), and on the basis of the strong resemblance with the genetic organization of IKe and Ff, we conclude that the open reading frames indicated code for these indispensable DNA replication functions. As in the phages IKe and Ff, in I2-2 the 3' terminal part of gene II

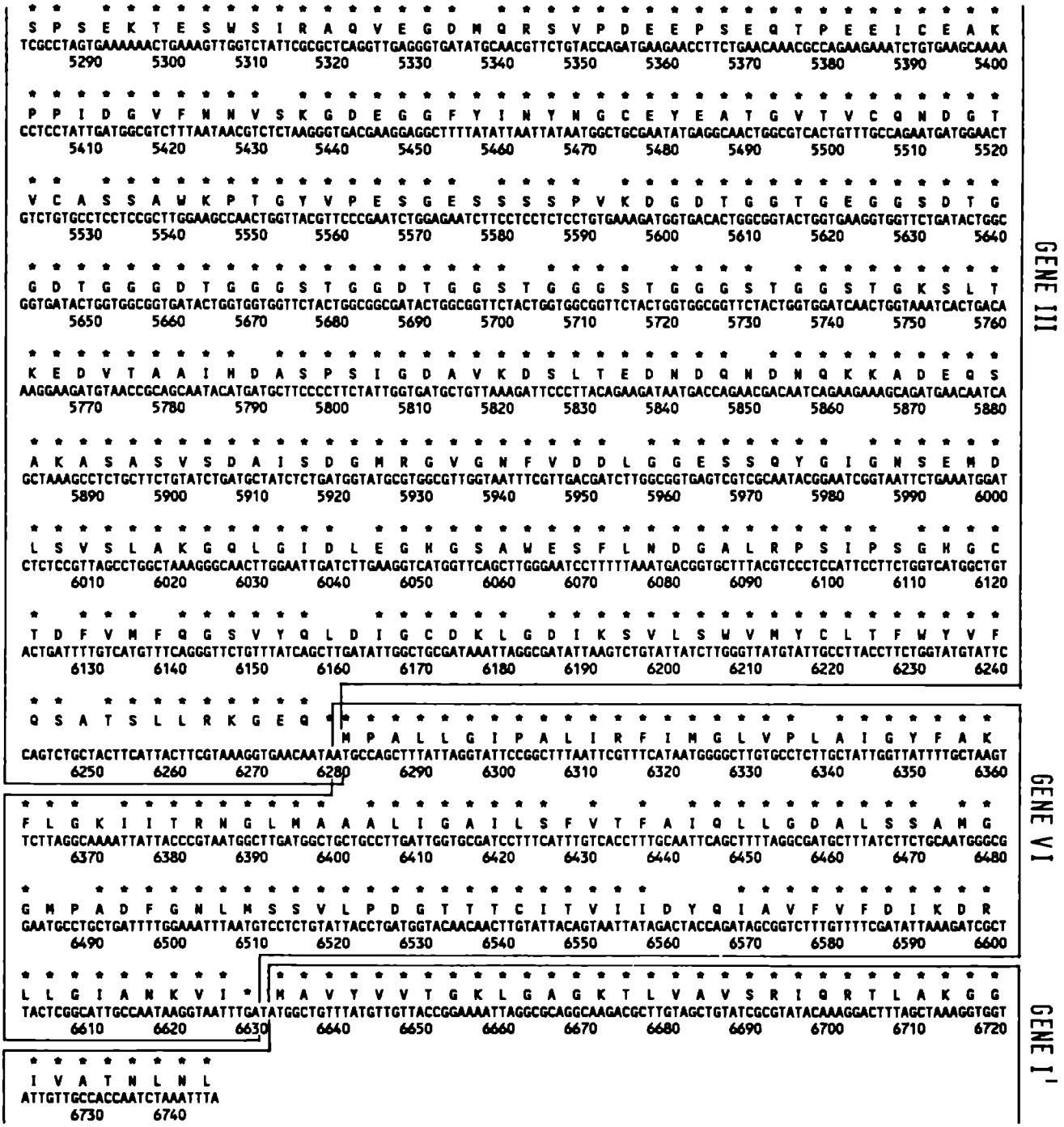


Figure 5.2 Continued

overlaps gene X completely. The proposed genetic organization is further substantiated by the observation that all initiation codons of the proposed genes and/or open reading frames are preceded by a sequence that is significantly homologous to the consensus sequence for ribosome binding (Table 5.2) (Shine and Dalgarno, 1974). All I2-2 genes start with an ATG codon, except gene III which starts with the codon GTG. Similar observations have been made during our studies of the genomes of Ff and IKe (van Wezenbeek *et al.*, 1980; Peeters *et al.*, 1985).

Genetic organization

The genetic maps of the genomes of I2-2, IKe and Ff (M13) are depicted in Figure 5.3 (van Wezenbeek *et al.*, 1980; Peeters *et al.*, 1985). The three genomes consist of an equal number of genes, which are arranged in an identical order (*vide supra*). Comparison

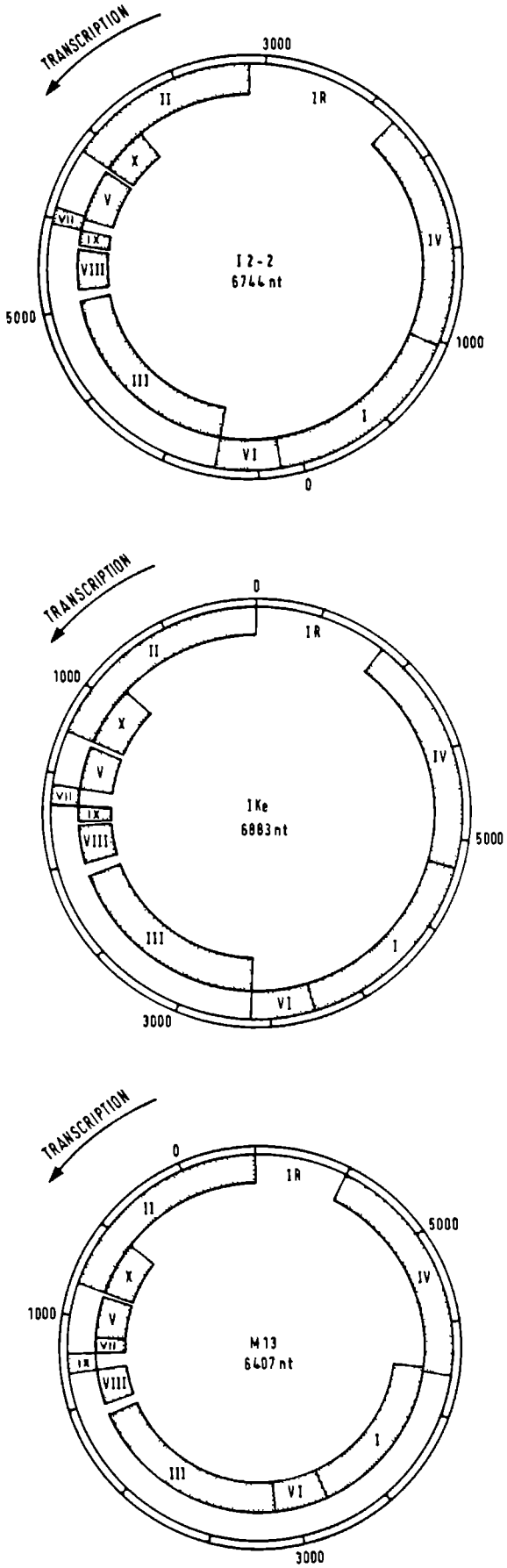


Figure 5.3 Comparison of the circular genetic maps of the genomes of the filamentous phages I2-2, IKe and Ff (M13).
Genes are indicated by Roman numerals. IR refers to the major intergenic region located between genes IV and II, in which both the origins for viral and complementary strand synthesis and the *cis*-acting element required for efficient phage assembly are located.

of the genomes of I2-2 and Ff reveals that, with the exception of genes II and VII, all I2-2 genes are larger than their Ff counterparts. The larger length of I2-2 compared with Ff is mainly the result of length differences between their respective major intergenic regions (Figure 5.3).

Apart from the major intergenic region that is located between genes IV and II, a much smaller one is located between genes VIII and III. It contains a ρ-independent transcription termination signal, identical in sequence to that of IKe and very similar to that of M13 (Figure 5.4) (van Wezenbeek *et al.*, 1980; Peeters *et al.*, 1985). In Ff and IKe the DNA region encompassing genes II through VIII is expressed in a nested set of mRNAs whose synthesis is initiated at different promoters but terminated at the afore mentioned ρ-independent transcription termination signal (Edens *et al.*, 1978). Because in the DNA region of I2-2 encompassing the major intergenic region and genes II to VIII, multiple nucleotide sequences are found which resemble the consensus sequence of *E. coli* promoters and which have an almost identical map position as the promoters found in IKe and Ff (data not shown), we suggest that this part of the I2-2 genome is expressed via a similar transcription mechanism.

Besides the complete overlap of gene X with the 3' terminal region of gene II, in I2-2 the reading frames of genes V and VII, VII and IX, and of III and VI overlap each

Table 5.2 Shine-Dalgarno sequences for the assigned and proposed genes of the I2-2 genome.
Nucleotides that are complementary to the 3'-terminal end of 16S ribosomal RNA are underlined (Shine and Dalgarno, 1974). Initiation codons are indicated in bold.

16S RNA	3'OH-AUUCCUCCACUAG
gene I	CAATAAGGTAATTTGAT.**ATG**
gene II	TAAGGGAATGCGT.**ATG**
gene III	TAAGGTGTAATT.**GTG**
gene IV	TGTGAAGGCTAATT.**ATG**
gene V	TAAGGGGTTTTAA.**ATG**
gene VI	CGTAAAGGTGAACAATA.**ATG**
gene VII	TGAAGAGCTGATTT.**ATG**
gene VIII	TAAGGAATACAC.**ATG**
gene IX	TGCTGGGGTACTCCG.**ATG**
gene X	TATGGGGCGTATCGAA.**ATG**

other with a few nucleotides. A similar compact genetic organization has found to be present in the genomes of IKe (Peeters *et al.*, 1985) and Ff (van Wezenbeek *et al.*, 1980), but with the exception that in Ff genes IX and VIII and genes I and IV overlap, rather than genes V and VII and genes III and VI (Figures 5.2 and 5.3).

Intergenic region

The major intergenic region (IR; 508 nucleotides) of phage Ff is a non-coding sequence that encompasses all *cis*-acting elements required for phage assembly and DNA replication (Zinder and Horiuchi, 1985). As one moves from the 5' to the 3' end, there are five consecutive regions (designated A through E) of dyad symmetry capable of forming stable stem-loop structures, followed by a AT-rich sequence without self-complementarity and that fulfils an enhancer function in the DNA replication process (Cleary and Ray, 1980; 1981; Dotto *et al.*, 1981; 1982). Region A functions both as a ρ-dependent transcription termination signal (Moses and Model, 1984; Smits *et al.*, 1984) and as a morphogenetic signal for the proper packaging of the viral strand into filamentous particles (Dotto and Zinder, 1983). Regions B and C function as origin for the synthesis of the complementary viral strand (Schaller *et al.*, 1976; Gray *et al.*, 1978), while the 'rabbit ears structure' consisting of regions D and E fulfils an indispensable role in the initiation and termination processes of viral strand replication (Meyer *et al.*, 1979).

The major intergenic region of the genome of IKe is 738 nucleotides long. Similar to Ff this intergenic region consists of five regions with dyad symmetry, which are homologous to those found in the intergenic region of Ff (Peeters *et al.*, 1985). The homology does not hold however for the major intergenic region of I2-2, although in this region of 817 nucleotides also five stretches with dyad symmetry are located and which resemble in number, relative position and structure those found in the intergenic regions of IKe and Ff (Figure 5.5). Only stretch A has a homologous counterpart in the genomes of Ff and IKe (Figure 5.5B). Since the phage encoded proteins that 'cross-talk' with this sequence (*i.e.* the coat proteins and probably also the phage assembly proteins encoded by genes I and IV), are highly conserved we propose that this sequence functions in I2-2 also as a phage morphogenesis signal. The observation that the secondary structure characteristics of the intergenic regions of I2-2, IKe and Ff are very much alike, highly favours the hypothesis that, despite the lack of homology, in I2-2 the DNA stretches of dyad symmetry have an identical function in the phage replication cycle as they have in the phages Ff and IKe.

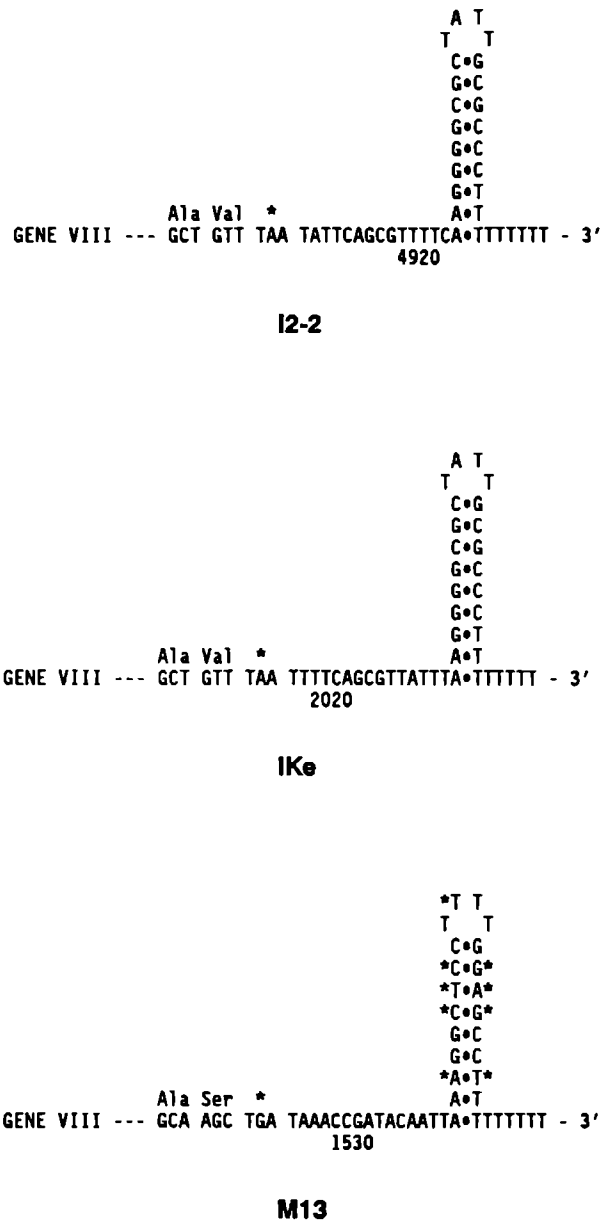


Figure 5.4 Nucleotide sequences and presumed secondary structures of the ρ-independent transcription termination signals of phages I2-2, IKe and Ff (M13), that are located in the intergenic region between genes VIII and III.
Nucleotides in the Ff sequence, encompassing the region of dyad symmetry, that are different from those in I2-2 and IKe are indicated with asterisks.

Evolution of the filamentous phage genome

Based upon the data presented it is unquestionable that the bacteriophages I2-2, IKe and Ff are evolutionarily related. Their genetic organization is almost identical and genes I, III, IV, VI, VII, VIII and IX, of I2-2 and IKe, are highly homologous. Furthermore there are no significant differences in codon usage and at the third codon position there is a strong preference for a T (46%). The evolutionary relationship between filamentous phages can be estimated by comparing the extent of nucleotide sequence similarity of genes I, III, IV, VI, VII, VIII and IX of I2-2, IKe and Ff (M13) respectively. The extent of synonymous and non-synonymous amino acid substitutions between the various genes was estimated according to Miyata and Yasunaga (1980) (Table 5.3). In virtually all comparisons

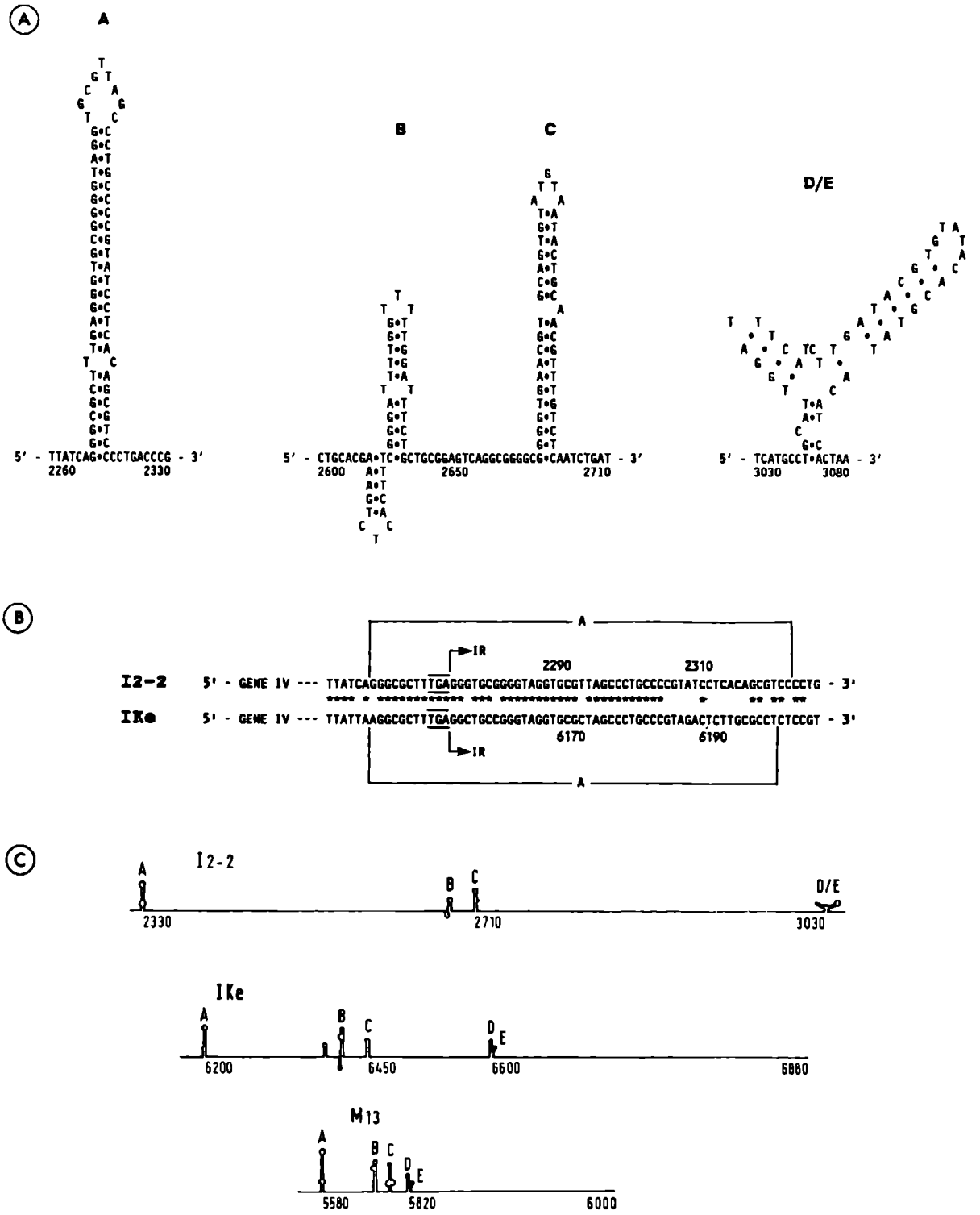


Figure 5.5 Comparison of the nucleotide sequence and secondary structural configuration of the regions of dyad symmetry in the major intergenic regions of the phages I2-2, IKe and Ff (M13).
(A) Nucleotide sequence (5'- 3') and potential secondary structures in the I2-2 major intergenic region.
(B) Comparison of the nucleotide sequence of region A of the intergenic region of the phages I2-2 and IKe. Within this sequence both the phage morphogenetic signal and a ρ-dependent transcription termination signal are located. Sequences are aligned in such a way that maximum homology is achieved. Regions of dyad symmetry are indicated by brackets. The stop codons of genes IV of I2-2 and IKe are over- and underlined.
(C) Comparison of 'morphology' and relative map position of the potential stem loop structures (hairpins A through E) in the major intergenic region of the phages I2-2, IKe and Ff (M13). The configuration are drawn to scale.

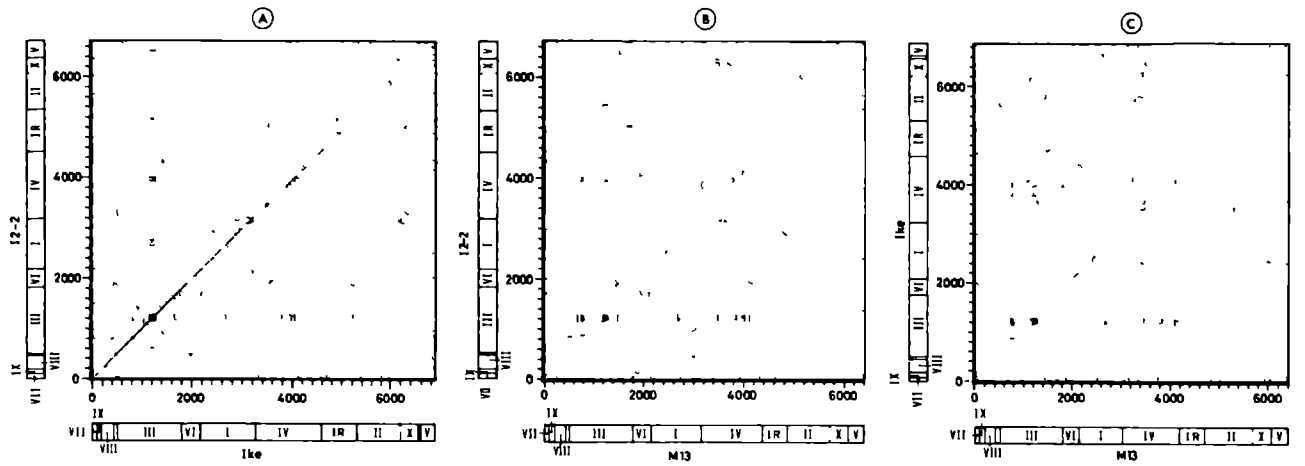


Figure 5.6 Dot-matrix analysis of the extent of homology between the genomes of the phages I2-2, IKe and Ff (M13).
Starting from the 5'-terminal nucleotide of gene VII consecutive stretches of 7 nucleotides are compared. Each dot represents a stretch of 7 identical nucleotides. Parallel to the axes the genetic maps are indicated.
(A) Dot-matrix of I2-2 *vs.* IKe. (B) Dot-matrix of I2-2 *vs.* Ff (M13). (C) Dot-matrix of IKe *vs.* Ff (M13).

he Ks value is larger than 1, indicating that the genes under consideration did not diverge recently. The reason why the nucleotide sequences of genes III and VIII of I2-2 and IKe are more strongly conserved than all others is unclear. The Ka values (Table 5.3), *i.e.* the values for the relative number of amino acid substitutions, clearly indicate that I2-2 is more closely related to IKe than to Ff (M13).

Genetic and cell biologic studies indicate that the protein encoded by gene III of Ff is specifically required for attachment of the phage to the conjugative pilus of the host cell. The minor, randomly distributed differences between the gene III proteins of IKe and I2-2 (Figure 5.2; 30 out of 434 amino acids), indicate that only a few amino acid substitutions are sufficient to alter/extend the plasmid specificity of the phage (*vide supra*). The observed serological relationship (Bradley *et al.*, 1983) between IKe, IKeh (a host range mutant of IKe) and I2-2, is in accordance with the high amino acid sequence homology found between the coat proteins encoded by gene VIII of both I2-2 and IKe. However, our sequencing data prove that, contrary to earlier speculations (Bradley *et al.*, 1983), I2-2 is not a host-range mutant of IKe. This can also be concluded from dot-matrix comparisons of the genomic sequences of I2-2, IKe and Ff (Figure 5.6). In these analyses the homology between successive stretches of heptanucleotides of the genomes of either I2-2 and IKe (Figure 5.6A), I2-2 and Ff (M13) (Figure 5.6B), or IKe and Ff (M13) (Figure 5.6C) are compared. The genomes are aligned by using the initiation codon of gene VII as a reference point. From these plots it can be concluded that genes VII, IX, VIII, III, VI, I and IV and the 5' end of the intergenic region (hairpin A) of I2-2 and IKe are highly similar. Immediately after the morphogenetic signal the sequence correspondence decreases abruptly. A similar observation has been

Table 5.3 Extent of synonymous- (Ks) and non-synonymous (Ka) substitutions in genes I through IX of the phages I2-2, IKe and Ff (M13).
The extent of synonymous and nonsynonymous substitutions was calculated as indicated in Materials and Methods.

	Sites Compared	Synonymous Difference (Ks)	Non-synonymous Difference (Ka)
I2-2 *vs.* IKe:			
gene I	365	>1	0.107
gene III	432	0.115	0.032
gene IV	427	>1	0.163
gene VI	116	>1	0.062
gene VII	29	>1	0.148
gene VIII	79	0.563	0.158
gene IX	33	>1	0.120
I2-2 *vs.* M13:			
gene I	342	>1	0.395
gene III	338	>1	0.709
gene IV	408	>1	0.458
gene VI	109	>1	0.737
gene VII	28	>1	0.573
gene VIII	63	>1	0.429
gene IX	31	>1	0.786
IKe *vs.* M13:			
gene I	342	>1	0.410
gene III	337	>1	>1
gene IV	409	>1	0.482
gene VI	109	>1	0.729
gene VII	31	>1	0.818
gene VIII	64	>1	0.571
gene IX	31	0.874	0.752

made when the genomes of I2-2 and Ff (M13) were compared, although in this case the similarity between the DNA sequence located between the 5' end of gene VII and the 3' end of hairpin A is less pronounced (Figure 5.6B).

The observation that no significant homology exists between the DNA regions of I2-2 and IKe or Ff that encompass the major part of the intergenic region (*vide supra*), genes II, X, and V suggests that the genome of I2-2 is composed of DNA sequences derived from two different replicons. We propose that the I2-2 genome is the result of a recombination event in which the replication cassette, *i.e.* the DNA fragment containing the viral and complementary strand origins of DNA replication and the phage encoded replication genes of its filamentous phage ancestor genome, has been exchanged for the 'rolling circle replication cassette' of another replicon. Whether the latter cassette originated from another filamentous phage or from a plasmid with a rolling circle replication mode is not yet clear, since computer searches in the EMBL (release 23) and Genbank (release 60) Nucleic Acids Databases has not yet revealed the existence of nucleotide sequences that are evolutionarily related to the replication module of I2-2.

CHAPTER 6

THE DNA BINDING WING OF THE SINGLE-STRANDED DNA BINDING PROTEINS OF THE FILAMENTOUS BACTERIOPHAGES M13 AND IKe: A RECURRENT MOTIF IN OTHER SINGLE-STRANDED DNA BINDING PROTEINS ?

ABSTRACT

^{1}H-NMR studies of single-stranded DNA binding proteins of the evolutionarily related filamentous *Escherichia coli* phages Ff and IKe have shown that these proteins contain a structural motif, the so-called DNA binding wing, that intimately is involved in their non-sequence-specific binding to single-stranded DNA. To obtain a first indication whether this DNA binding wing is a recurrent motif in other single-stranded DNA binding proteins, we have investigated whether the amino acid sequences of these wings are recurrent sequences in other known prokaryotic and eukaryotic single-stranded DNA binding proteins. The analyses revealed that of all ssDNA binding proteins whose amino acid sequence has been stored in the protein databases, only ssDNA binding proteins encoded by the bacteriophages Ff, IKe, Pf1, Pf3, ϕ29, If1, I2-2, and T4 plus the eukaryotic intermediate filament protein vimentin, comprise an amino acid sequence that exhibits a significant degree of homology with this sequence.

INTRODUCTION

Single-stranded DNA binding proteins (ssDBPs) are ubiquitous in nature. They are produced both by prokaryotic and eukaryotic cells and a large number of viral genomes code for a ssDBP as well. All ssDBPs identified so far perform a wide variety of functions, particularly in the processes of DNA replication, repair, and recombination, and mRNA translation (Kornberg and Baker, 1992). The proteins of which the physico-chemical properties have been established, have shown to bind to single-stranded DNA (ssDNA) aspecifically yet with high affinity and co-operativity (Kornberg and Baker, 1992).

Examples of prokaryotic ssDBPs are the proteins encoded by gene V (gVp) of the evolutionarily related filamentous *Escherichia coli* phages Ff (M13, fd, and f1), IKe, and I2-2 (Chapter 5; Alberts *et al.*, 1972; Oey and Knippers, 1972; Peeters *et al.*, 1983) and gene 32 protein (g32p) encoded by the *E. coli* phage T4 (Alberts and Frey, 1970). Other prokaryotic ssDBPs are the *E. coli* encoded single-stranded DNA binding protein SSB (Sancar *et al.*, 1981) and the RecA protein (Clark and Margulies, 1965). Examples of eukaryotic ssDBPs are mtSSB of *Xenopus laevis* (Mahoungou *et al.*, 1988; Ghrir *et al.*, 1991; Tiranti *et al.*, 1991), p16 and SSB encoded by rat mitochondrial DNA (Cobianchi *et al.*, 1986; Hoke *et al.*, 1990), and RIM1 and SSB encoded by *Saccharomyces cerevisiae* (Heyer *et al.*, 1990; van Dyck *et al.*, 1992).

Of all these proteins, the gVps of the phages Ff and IKe, the g32p of phage T4 and RecA protein are studied best (for reviews, see Chase and Williams, 1986; Model and Russel, 1988; Roca and Cox, 1990). By regulating not only the production of progeny viral strands, but also, at the level of translation, the synthesis of a number of phage specific proteins, gVp fulfils an indispensable function in the phage replication process (Zaman *et al.*, 1990; 1991). By protecting ssDNA intermediates from nuclease attack in processes such as DNA replication, repair and recombination, g32p plays a key role in the replication of phage T4 DNA. Translation studies have furthermore demonstrated that it autoregulates its synthesis via specific binding to particular sequences in its cognate mRNAs (Chase and Williams, 1986).

A first report on the three-dimensional crystal structure of gVp of phage Ff has already been reported a decade ago (Brayer and McPherson, 1983). Recent 2D-NMR studies have, however, indicated that the structure as published is not correct and that it deviates significantly from that present in solution (van Duynhoven *et al.*, 1990; Folkers *et al.*, 1991; 1993b). In both analyses it was found that the protein almost entirely consists of β-sheets. The structures differ, however, significantly both in the registers of amino acid residues in particular β-sheets and the three-dimensional folding of the polypeptide chain (Folkers *et al.*, 1993a; 1994). Interestingly, renewed crystallographic analyses of gVp have very recently confirmed the conclusion that the crystal structure as proposed originally is not correct and that the 3D-structure of gVp

as present in the crystal is almost identical to that found in solution with the aid of 2D and 3D-NMR analyses (Folkers *et al.*, 1991; Skinner *et al.*, 1994; Folkers *et al.*, 1994). With the aid of 2D-NMR the secondary structure of gVp of phage IKe has also been elucidated. Comparison of this structure with that of Ff revealed that they are almost identical, notwithstanding the fact that the amino acid sequence identity between the gVps of Ff and IKe is only 45% (de Jong *et al.*, 1989b; van Duynhoven *et al.*, 1992).

2D-NMR analyses have also been used to map, with the aid of non- and spin-labelled oligonucleotides, the DNA binding domains in gVp of Ff and IKe (Chapter 4; de Jong *et al.*, 1989b; van Duynhoven *et al.*, 1993; Folkers *et al.*, 1993b). From these studies it could be concluded that, in contrast to the proposals made previously (Brayer and McPherson, 1984a; Hutchinson *et al.*, 1990), the DNA binding site is situated inside instead of outside of the protein and that it is composed of amino acid stretches derived from each of the subunits of the protein dimer (de Jong *et al.*, 1989b; van Duynhoven *et al.*, 1993; Folkers *et al.*, 1993b). One of the stretches consists of the so-called DNA-binding wing (comprising amino acid residues Arg16-Glu30 in gVp of Ff (van Duynhoven *et al.*, 1990; 1993; Folkers *et al.*, 1993b) and residues Arg16-Glu31 in gVp of IKe (de Jong *et al.*, 1989b; van Duynhoven, 1993). The other two stretches are part of the so-called complex-loop (residues Leu44-Asp50 in gVp of Ff and Asn47-Glu51 in gVp of IKe) (Brayer and McPherson, 1983) and dyad-loop (residues Lys69-Arg80 in gVp of Ff and Lys70-Arg81 in gVp of IKe), respectively. The proton resonances of the amino acid residues in the DNA binding wing are affected most by the spin-labelled oligonucleotides. The importance of this structural element in ssDNA binding is additionally strengthened by the fact that the residues constituting this element are highly conserved among the gVps of Ff and IKe, whereas the amino acid sequences of the other stretches are clearly more diverged (Peeters *et al.*, 1983; Luiten *et al.*, 1985; de Jong *et al.*, 1989b).

Now that detailed information is available both of the structure of gVp and of the role/function of the DNA binding wing in ssDNA binding it is of interest to investigate whether, in analogy to the DNA binding motifs of dsDBPs (Pabo and Sauer, 1992), the DNA binding wing is a recurrent DNA binding motif in other ssDBPs.

As structural information of ssDBPs other than the gVp is not (yet) available, similarities in amino acid sequences were investigated with the aid of computer assisted structural analyses.

MATERIALS AND METHODS

Proteins

The prokaryotic and eukaryotic single-stranded DNA binding proteins found to be present in the three protein databases PIR1 (annotated and classified entries), PIR2 (preliminary entries), and PIR3 (unverified entries) are: gVp encoded by filamentous bacteriophages Ff (Cuypers *et al.*, 1974; Nakashima *et al.*, 1974), IKe (Peeters *et al.*, 1983), I2-2 (Chapter 5), Pf1 (Maeda *et al.*, 1982), Pf3 (Putterman *et al.*, 1984; Luiten *et al.*, 1985), and If1 (Carne *et al.*, 1991); bacteriophage ϕ29 p5 (Yoshikawa and Ito, 1982; Martín and Salas, 1988); bacteriophage T4 g32p (Williams *et al.*, 1981; Krisch and Allet, 1982); bacteriophages T2, T3, T6, and T7 DBP (Dunn and Studier, 1981; Schmitt *et al.*, 1987; McPheeters *et al.*, 1988; Beck *et al.*, 1989); *E. coli* RecA protein (Horii *et al.*, 1980; Sancar *et al.*, 1980), SSB (Sancar *et al.*, 1981), and SSF (Chase *et al.*, 1983); adenovirus type 2 (Kruijer *et al.*, 1982), 4 (Kitchingman, 1985), 5 (Kruijer *et al.*, 1981), 7 (Quinn and Kitchingman, 1984), and 12 DBP (Kruijer *et al.*, 1983); herpes simplex virus type 1 ICP8 (Quinn and McGeoch, 1985); varicella-zoster virus DBP (Davison and Scott, 1986); Epstein-Barr virus DBP (Baer *et al.*, 1984); *Xenopus laevis* mtSSB (Mahoungou *et al.*, 1988; Ghrir *et al.*, 1991; Tiranti *et al.*, 1991); cytomegalovirus DB129 (Anders, 1990); bovine leukemia virus p12, and bovine protein UP1 (Copeland *et al.*, 1983; Merrill *et al.*, 1987); Rauscher murine leukemia virus p10 (Henderson *et al.*, 1981); rat mitochondrial p16 and SSB (Cobianchi *et al.*, 1986; Hoke *et al.*, 1990); *Saccharomyces cerevisiae* RIM1, and SSB (Heyer *et al.*, 1990; van Dyck *et al.*, 1992); DBP encoded by plasmids ColIb-P9 (Howland *et al.*, 1989), R64 (Ruvolo *et al.*, 1991), pIP71a (Ruvolo *et al.*, 1991), and pIP231a (Ruvolo *et al.*, 1991); *Drosophila melanogaster* SSB (Haynes *et al.*, 1987); human heterogeneous ribonuclear particle protein A1, and replication protein A (Buvoli *et al.*, 1988; Biamonti *et al.*, 1989; Erdile *et al.*, 1991). The vimentins encoded by man (Ferrari *et al.*, 1986) and chicken (*Gallus gallus*) (Zehner *et al.*, 1987) have also been included in this study.

Amino acid sequence comparisons

Amino acid sequence comparisons were performed using the computer programme BESTFIT, which is part of the Sequence Analysis Software Package (Version 7.1) of the Genetics Computer Group (University of Wisconsin, Madison)(Devereux *et al.*, 1984). The BESTFIT programme enables an optimal local alignment of related amino acid sequences. This optimal alignment is found via the insertion of gaps to maximize the number of matches using the algorithm of Smith and Waterman (1981). To calculate the best alignment, BESTFIT uses a symbol comparison table, which has been derived by Gribskov and Burgess (1986) from the mutational difference matrix (MDM78) of Schwartz and Dayhoff (1978). The number and size of gaps is restricted by the use of a gap penalty. The penalty for each gap is 3 and for every residue in a gap 0.1. As a result the penalty for the smallest possible gap is 3.1. This value, which is two times higher than the largest score in the symbol comparison table (being 1.5) ensures that gaps are only inserted when they significantly improve the alignment.

Evaluation of the statistical significance of the alignments

The amino acid sequence alignments found by the BESTFIT programme can reflect sequence similarity

that is due to common ancestry or simply a locally biased amino acid composition. To test whether the alignment of two amino acid sequences indeed reflects a statistically significant relationship, the computer programme RDF2G was used (Lipman and Pearson, 1985; Pearson and Lipman, 1988; Pearson, 1990). This programme calculates the statistical significance of an alignment in a three step procedure. In the first step, the so-called similarity score (gap-penalty = 0) of the optimal alignment of two amino acid sequences is calculated, using the PAM250 scoring matrix of Dayhoff, Schwartz, and Orcutt (1978). In the second step of the procedure, one of the two aligned amino acid sequences is shuffled 200 times, resulting in randomized sequences that have the same lengths and amino acid compositions as the original sequence. These 200 randomized sequences are aligned with the unshuffled sequence and the so-called random scores, the mean, and standard deviation of these random scores are calculated using the PAM250 scoring matrix. Subsequenty, in the third step, the statistical significance of the optimal alignment expressed as the *z*-value is calculated, where

$$z = \frac{(\text{similarity score} - \text{mean of random scores})}{(\text{standard deviationof random scores})}$$

In accordance with Lipman and Pearson (1985; Pearson, 1990), a *z*-value of 3 or greater is considered to be required to permit, with reasonable confidence, the conclusion that the alignment of the two sequences is statistically significant. This arbitrary threshold value is supported by *z*-values calculated for evolutionarily related proteins (Lipman and Pearson, 1985; Pearson, 1990).

RESULTS

Amino acid sequence comparisons

The amino acid sequence from residue 16 to 29 of gVp of IKe, and the slightly different amino acid sequence from residue 16 to 28 of gVp of Ff, were used to search for homologous amino acid sequences in the ssDBPs listed in Materials and Methods. The analyses revealed that both sequences gave the same optimal alignments in the proteins under investigation. The proteins that contain such a homologous amino acid sequence are listed in Table 6.1. From this table it can be concluded that several bacteriophage encoded ssDBPs show significant homology. Table 6.1 also includes the intermediate filament protein vimentin of an eukaryotic species (human). This protein was included in our analyses because recent structural and functional analyses of the N-terminal domain of vimentin (Traub *et al.*, 1992) have strongly indicated that it might contain an amino acid sequence highly homologous to that of the DNA binding wing of gVp.

The amino acid sequences of the proteins that gave the best alignments are depicted in Figure 6.1. Amino acid residues that are identical to that in the DNA binding wing of IKe are boxed by double lines, whereas conservative amino acid substitutions are indicated by single boxes. From this figure, it can be concluded that, as one would expect, the number of identical/conserved residues shared with the amino acid sequence of the DNA binding wing of IKe, varies considerably. The I2-2 sequence shares only 5 residues, whereas the ssDBP of phage Pf3 shares 9 out of 14 amino acid residues.

Table 6.1 ssDNA binding proteins which contain an amino acid sequence homologous to the ssDNA binding wing of gVp of phages IKe and Ff.

Protein	Phage/Species
DBP	*Pseudomonas aeruginosa* phage Pf3
p5	*Bacillus subtilis* phage φ29
gVp	*Pseudomonas aeruginosa* phage Pf1
gVp	*Escherichia coli* phage I2-2
DBP	*Escherichia coli* phage If1
g32p	*Escherichia coli* phage T4
vimentin	human

Statistical significance of the alignments

To see whether the alignments depicted in Figure 6.1 are statistically significant, the statistical significance of these alignments was calculated. In these calculations the PAM 250 scoring matrix of Dayhoff, Schwartz, and Orcutt (1978) was used. All the amino acid sequences depicted in Figures 6.1 and 6.3 have been mutually compared and the statistical significances expressed as *z*-values calculated. The results are listed in Table 6.2. The *z*-values have been calculated for the sequences comprising the (putative) DNA binding wings (a), as well as for the same sequences extended by two residues at the C-terminal end (b). The latter calculations have been performed because the glutamic acid residue at position 31 in IKe (and at position 30 in Ff) showed enhanced relaxation in the presence of spin-labelled oligonucleotides (de Jong *et al.*, 1989b; van Duynhoven *et al.*, 1993; Folkers *et al.*, 1993b).

Evaluation of the *z*-values requires a discrimination between statistically significant and not significant values. It has been argued previously (Lipman and Pearson, 1985; Gribskov and Burgess, 1986; Pearson, 1990) that one should be skeptical of conclusions based on sequence similarity scores with *z*-values less than 3, and more confident when the *z*-values are greater than 6 (this agrees with our results; *vide infra*). Comparison of the 'DNA binding wings' of the ssDBPs encoded by phages IKe, Ff and Pf3 shows that the high level of identity is expressed in a *z*-value greater than 6: IKe *vs.* Ff (z=7.5), IKe *vs.* Pf3 (z=6.8), and Ff *vs.* Pf3 (z=6.6). Less well conserved amino acid sequences, show *z*-values between 3 and 6: for instance, Ff *vs.* human vimentin (z=4.0), I2-2 *vs.* Pf3 (z=3.5), and If1 *vs.* Ff (z=3.1). Finally, a number of amino acid sequence comparisons did not show a significant alignment (Table 6.2, z<3).

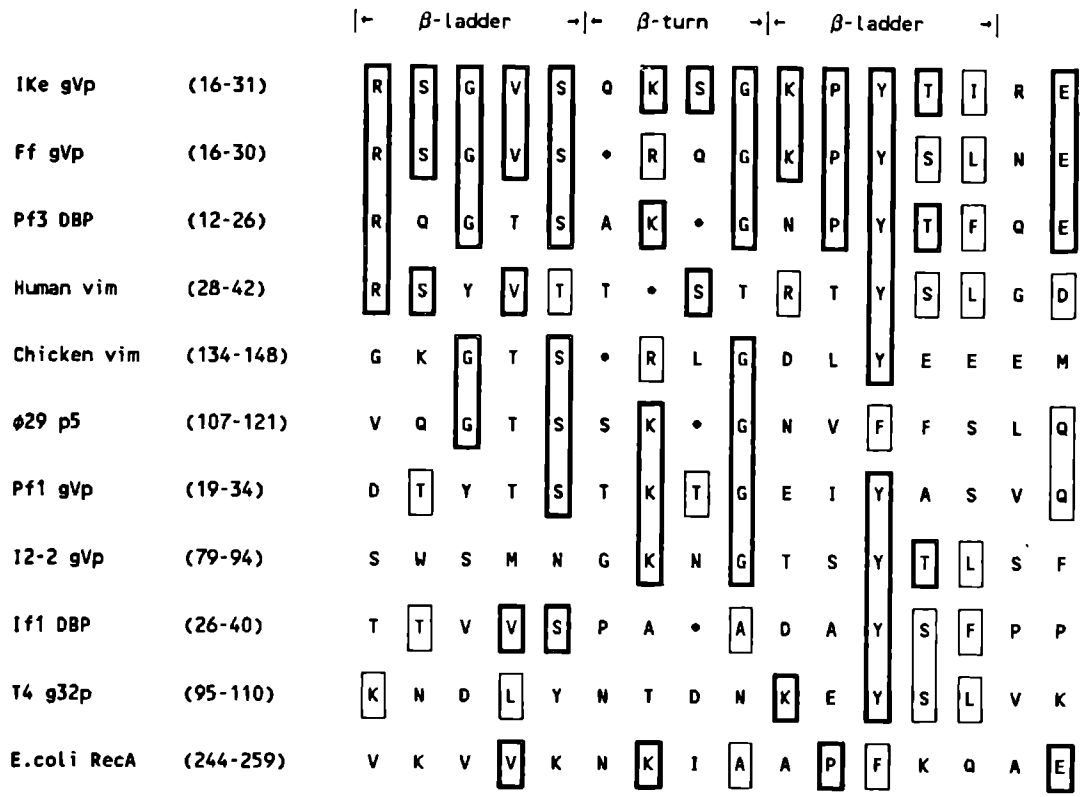


Figure 6.1 Alignments of the amino acid sequences comprising the DNA binding wings of phages Ff and IKe gVp, and the putative DNA binding β-loop structures of other DBPs.
The similarities between the IKe amino acid sequence and the sequences of the other DBPs, are indicated by boxes. Residues identical to the IKe gVp sequence are boxed by double lines, whereas conservative replacements are boxed by single lines. The dots represent gaps inserted in the amino acid sequences to obtain optimal alignments.
The amino acids are represented by one-letter symbols.

DISCUSSION

General remarks

Although molecular biological, and biophysical studies have delivered a wealth of information about the biological, and physico-chemical properties of ssDBPs, there is a scarcity of information regarding the domains of the proteins involved in binding to ssDNA. In an attempt to tentatively map these sites, amino acid sequence comparisons of ssDBPs are the only tools currently available to search for similarities.

Prasad and Chiu (1987) were the first to compare the amino acid sequences of several ssDBPs. They used the computer programmes RELATE and ALIGN (Dayhoff *et al.*, 1983), which both are based on the algorithm of Needleman and Wunsch (1970), to search for similarities between the Ff gVp amino acid sequence and the amino acid sequences of other ssDBPs. Since the proteins they compared (Ff gVp, IKe gVp, T4 g32p, *E. coli* RecA, *E. coli* SSB, and *E. coli* SSF) lack for the most part strong overall sequence homology, they used the entire Ff gVp sequence as a query sequence and searched for domains that yielded the best alignment with the positively charged and aromatic amino acid residues. Although the domains they identified, *i.e.* residues 1-88 of IKe gVp, residues 1-87 of Ff gVp, residues 32-113 of T4 g32p, residues 225-312 of *E. coli* RecA, residues 29-117 of *E. coli* SSB, and residues 29-115 of *E. coli* SSF, show alignment of four charged, and five aromatic amino acid residues, they were unable to pinpoint discrete regions in these proteins which may be involved in ssDNA binding (Prasad and Chiu, 1987). This limited success of the amino acid sequence comparisons using the complete protein sequences of several DBPs, is due to the low overall homology of the proteins investigated. The increased knowledge about the three-dimensional structure and ssDNA binding domains of gVp of the phages Ff and IKe, has for two reasons stimulated us to repeat the computer assisted analyses, but now in a completely different way. Firstly, the gVps of the phages IKe and Ff share a structural motif, comprising amino acid residues Arg16 to Ile29 in IKe and Arg16 to Leu28 in Ff respectively, that is intimately involved in DNA binding (Chapter 4; King and Coleman, 1987; de Jong *et al.*, 1989b; van Duynhoven *et al.*, 1990; 1993; Folkers *et al.*, 1993b). Secondly, the amino acid sequences comprising these DNA binding wings are very similar (the overall sequence identity between gVp of Ff and IKe is only 45%). Hence, it is of interest to inves-

Table 6.2 Statistical significance expressed as *z*-values, of the amino acid sequence alignments depicted in Figure 6.1.
The amino acid sequences compared are (*Cf.* Figure 6.1): IKe gVp residues 16-29 (a) and 16-31 (b), Ff gVp residues 16-28 (a) and 16-30 (b), Pf3 DBP residues 12-24 (a) and 12-26 (b), human vimentin residues 28-40 (a) and 28-42 (b), chicken vimentin residues 134-146 (a) and 134-148 (b), φ29 p5 residues 107-119 (a) and 107-121 (b), Pf1 gVp residues 19-32 (a) and 19-34 (b), I2-2 gVp residues 79-92 (a) and 79-94 (b), If1 DBP residues 26-38 (a) and 26-40 (b), T4 g32p residues 95-108 (a) and 95-110 (b), and *E. coli* RecA residues 244-257 (a) and 244-259 (b).

	E.coli RecA		T4 g32p		If1 DBP		I2-2 gVp		Pf1 gVp		φ29 p5		chicken vimentin		human vimentin		Pf3 DBP		Ff gVp	
	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b
IKe gVp	1.8	2.2	3.5	2.5	2.4	2.0	4.4	3.6	0.9	0.9	1.6	1.7	1.2	0.6	3.6	3.1	6.8	7.6	7.5	8.3
Ff gVp	1.5	2.3	4.0	3.1	3.1	2.2	4.3	3.6	0.5	0.3	1.3	< 0	1.5	1.3	4.0	4.4	6.6	7.3		
Pf3 DBP	1.1	1.9	< 0	0.5	2.7	2.6	3.5	1.8	1.0	1.0	3.4	3.5	2.8	2.5	< 0	< 0				
hum vim	< 0	0.5	2.0	1.6	1.7	1.3	1.1	< 0	2.0	1.9	< 0	< 0	< 0	< 0						
chi vim	3.1	2.0	< 0	< 0	0.1	< 0	< 0	< 0	2.7	2.1	3.6	2.5								
φ29 p5	1.2	1.1	< 0	< 0	< 0	< 0	< 0	< 0	1.9	3.5										
Pf1 gVp	< 0	< 0	< 0	< 0	0.3	0.2	< 0	< 0												
I2-2 gVp	1.7	0.0	1.7	0.5	0.8	< 0														
If1 DBP	0.9	0.6	< 0	< 0																
T4 g32p	< 0	< 0																		

igate whether this motif is also shared by other ssDBPs of prokaryotic and eukaryotic origin.

Amino acid sequence alignments

One of the advantages of studying similarities between amino acid sequences by computer is the ability o find expected as well as unexpected sequence nomologies that are statistically significant. However, if here is little or no information available regarding the unctional domains of the proteins, the significance of he observed homology can be quite puzzling.

Our computer analysis shows that only some phage encoded ssDBPs contain amino acid sequences that are significantly homologous with the ssDNA binding wings of the gVps of the phages Ff and Ke. The eukaryotic intermediate filament protein imentin has been included in the analyses for a completely different reason (*vide supra*). As shown, his protein also contains in its N-terminal end amino cid sequences that are significantly homologous to hat of the DNA binding wings. The alignments found by our computer analysis include sequences (*e.g.* Pf3 DBP) that already were expected to be present on the basis of other investigations (de Jong *et al.*, 1989b). The observed amino acid sequence homology between Pf1 gVp and the DNA binding wing is also consistent with the observations made by Plyte and Kneale (1991). The statistical significance of the observed homology is, however, very low (*vide infra*). Apart from the homologies that already have been observed before, we have also observed homologies, *e.g.* in g32p, that are different from those proposed in earlier studies (de Jong *et al.*, 1989b). An interesting observa-tion was made concerning the identified sequence in protein p5 of phage φ29. If this sequence is compared with the identified sequence in the ssDBP of phage Pf3, and with the DNA binding wings in IKe and Ff, it displays the highest statistical significance with the former one (see Figure 6.1).

Comparison of the (putative) DNA binding wings

On the basis of the secondary structure elucidated for the DNA binding wings of the gVps of Ff and IKe (de Jong *et al.*, 1989b; van Duynhoven *et al.*, 1990), the homologous amino acid sequences selected by the computer programme in the other ssDBPs (Figure 6.1) have been folded in a similar β-ladder conformation (Figure 6.2).

From a comparison of these folds several properties/characteristics that are shared by the majority of these sequences became apparent. Firstly, all β-ladder structures contain at a similar/homologous position one of the two aromatic amino acid residues tyrosine or phenylalanine, which in gVp of Ff (Tyr26) and IKe (Tyr27) have shown to be intimately involved in DNA binding (Chapter 4; de Jong *et al.*, 1989b; van Duynhoven *et al.*, 1993). Secondly, nine out of the eleven structures contain at the C-terminal end of the β-turn one of the two smallest amino acid residues, *i.e.* glycine or alanine. Thirdly, in 60% of the β-ladders a serine residue immediately precedes the β-turn.

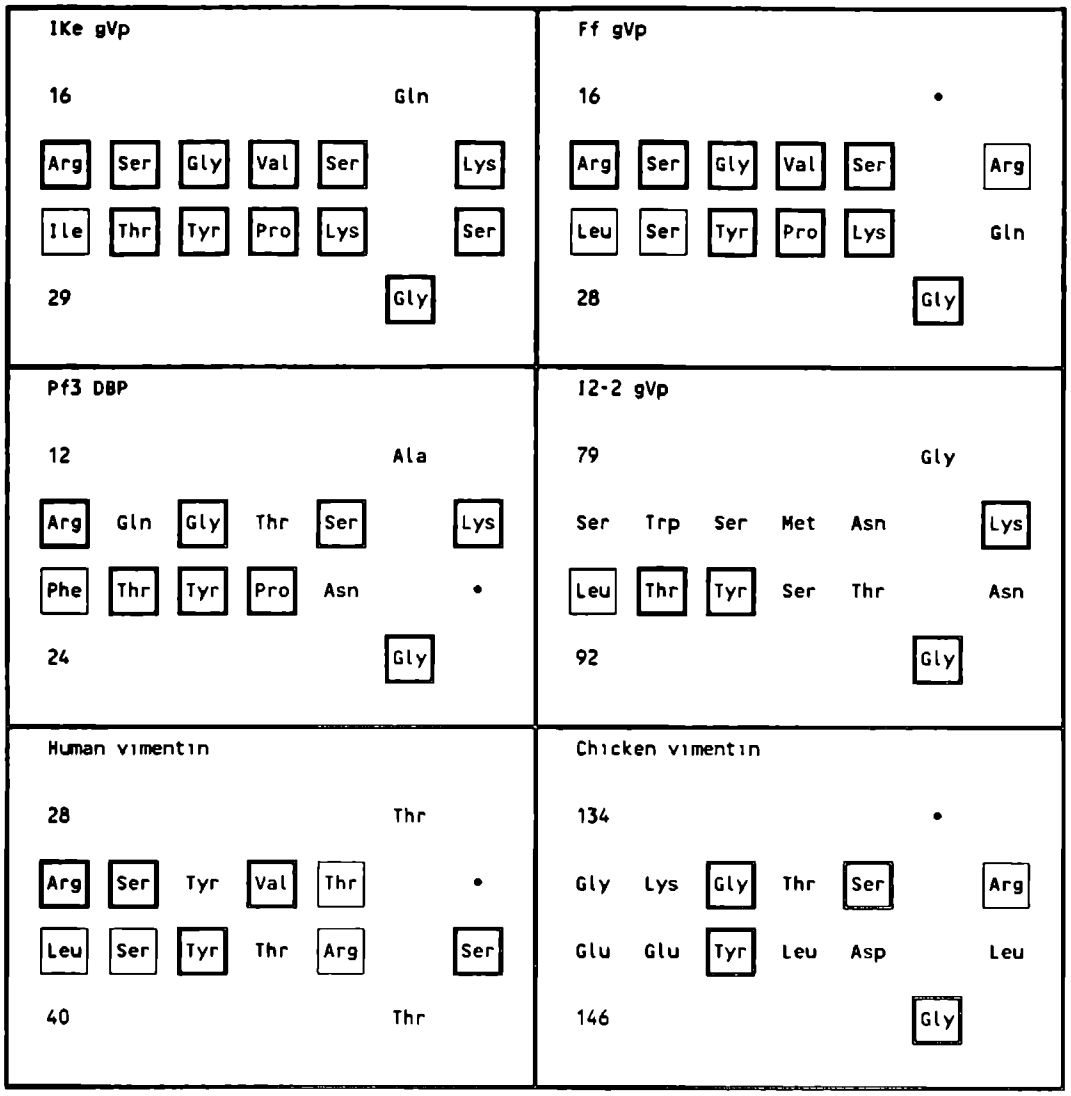


Figure 6.2 Schematic drawings of the DNA binding wings of gVp of the phages Ff and IKe, and of the putative DNA binding wings of the ssDNA binding proteins of the phages Pf1, Pf3, I2-2, ϕ29, T4, If1, and of *E. coli* RecA, and human and chicken encoded vimentin.
Residues identical to the IKe gVp sequence are boxed by double lines, whereas conservative replacements are boxed by single lines The dots represent gaps inserted in the amino acid sequences to obtain optimal alignments

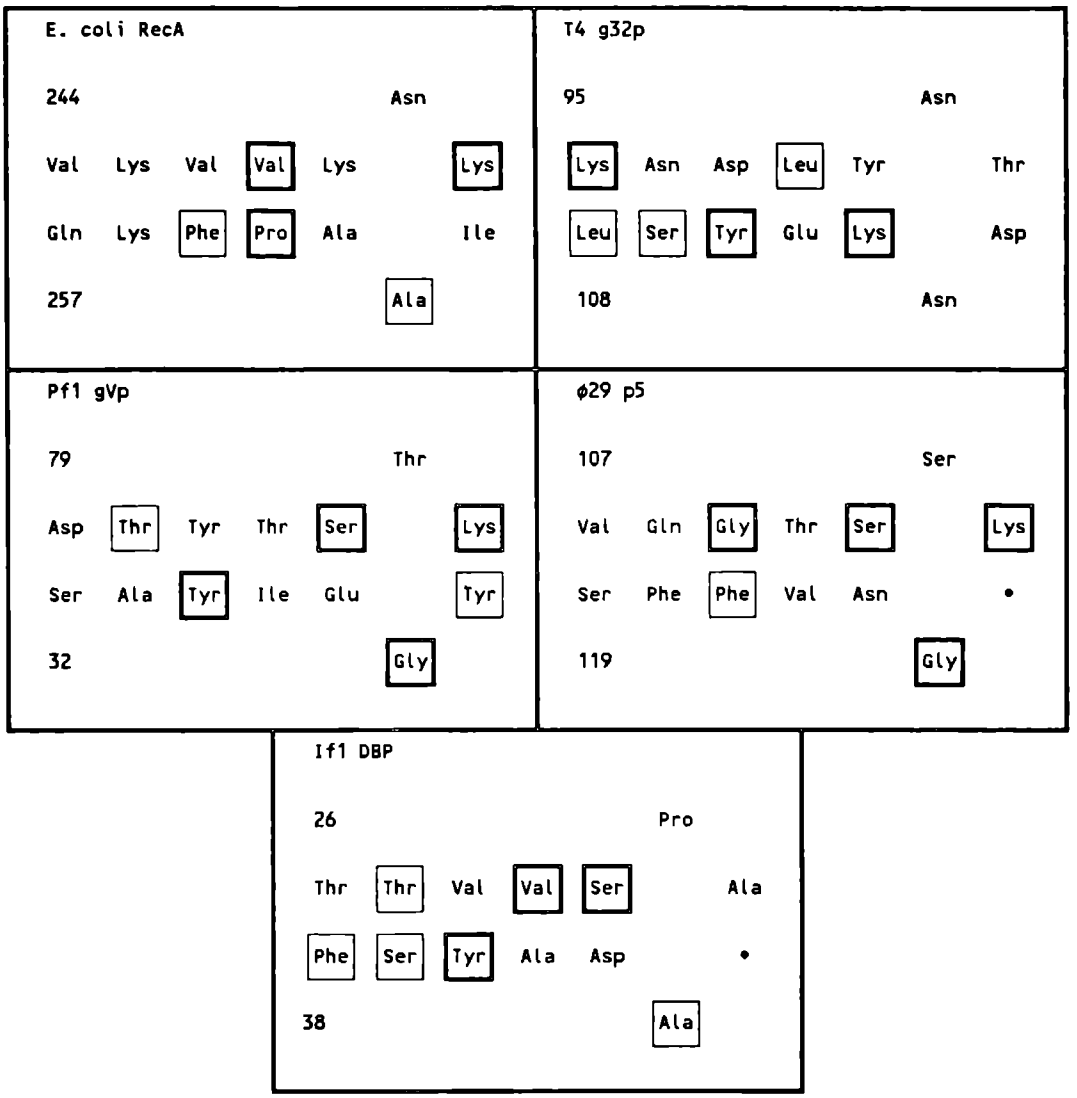
Interestingly enough the occurrence of this amino acid at this position almost parallels the occurrence of the glycine residue at the end of the β-turn and in the middle of the first β-strand Finally two-third of the structures contain a basic amino acid residue in the N-terminal half of the β-turn These residues are homologous to the basic residues in gVp of Ff (Arg21) and IKe (Lys22) that have been demonstrated to be intimately involved in ssDNA binding (van Duynhoven *et al*, 1993, Folkers *et al.*, 1993b)

Biological significance of the observed similarities

In the subsequent paragraphs the biological significance of the results of our computer analyses in connection with our knowledge of the ssDNA binding properties of the respective proteins will be discussed.

Pf3 DBP, Pf1 gVp and ϕ29 p5

The high statistical significance of the alignments of residues 12-24 of Pf3 DBP and the amino acid sequences comprising the DNA binding wings of gVp of Ff and IKe highly supports the hypothesis made previously (de Jong *et al*, 1989b) that Pf3 DBP contains a DNA binding motif whose three dimensional structure is similar to that of the DNA binding wing of Ff and IKe As a matter of fact preliminary 2D-NMR analyses on wild-type and mutant Pf3 DBP have recently indicated that indeed this conclusion seems to be correct (R.H.A. Folmer, personal communication).



In contrast to the ssDBP encoded by Pf3, the gVp encoded by phage Pf1 does not contain an amino acid sequence with such a high homology to the DNA binding wing. Despite of this, there are several indications that support the hypothesis that the sequence selected indeed is involved in ssDNA binding (Plyte and Kneale, 1991). Firstly, with the aid of chemical modification experiments it has been demonstrated that in the free Pf1 protein Lys25 and Tyr30 are surface accessible residues, whereas they are protected from modification in the nucleoprotein complex with phage DNA (Tsugita and Kneale, 1985; Plyte and Kneale, 1991). Secondly, secondary structure predictions by the method of Chou and Fasman, suggest that the region encompassing residues 17 through 34 will adopt a β-sheet conformation (Chou and Fasman, 1978; Maeda *et al.*, 1982; Plyte and Kneale, 1991).

The protein encoded by gene 5 of the *Bacillus subtilis* phage ϕ29 is indispensable for phage DNA replication. It is required *in vivo* in the elongation of ϕ29 DNA replication, and *in vitro* studies have indicated that the purified protein binds to ssDNA and protects it against nuclease degradation (Salas, 1988; Martín *et al.*, 1989). Which domain(s) of the protein is (are) involved in ssDNA binding is not yet known. Interestingly, the putative DNA binding domain (residues 107-119), as selected by the computer search programme, has a significantly higher degree of homology with the putative ssDNA binding wing of phage Pf3 than with the DNA binding motif of gVp of Ff or IKe (Table 6.2; *Cf.* Figure 6.3). These observations as well as the high degree of homology between the (putative) ssDNA binding wings of Pf3, IKe, and Ff previously mentioned, lend support to the hypothesis that the ssDBPs of Pf3, ϕ29, IKe, and Ff are evolutionarily related.

T4 g32p

The product of gene 32 of bacteriophage T4 (g32p) was the first ssDBP of which the ssDNA binding properties were characterized *in vitro* (Chase and Williams, 1986). Prasad and Chiu (1987) have demonstrated that the sequence running from residues 32 to 113 shows significant homology with that of gVp of phage Ff. Furthermore, ^{1}H-NMR experiments involving

complexes of g32p with perdeuterated tyrosine and phenylalanine residues, and $d(pA)_n$ (n=2,4,6,8, and 10) or $d(pT)_8$, have demonstrated that the amino acid sequence running from Tyr73 to Tyr115 forms (part of) the DNA binding domain, and that six aromatic residues, and the tyrosine residue at position 115 in particular, are intimately involved in the ssDNA binding (Prigodich *et al.*, 1984; 1986; Casas-Finet *et al.*, 1988; Pan *et al.*, 1989; Shamoo *et al.*, 1989).

Our analysis shows that the amino acid sequence running from residue 95 to 108 in g32p is statistically significantly homologous to the amino acid sequences of the DNA binding wings in gVp of Ff and IKe (z=4.0 and z=3.5, respectively (Table 6.2)). Because this sequence is located within the experimentally determined amino acid stretch that has shown to be involved in ssDNA binding, it is most likely that in binding of g32p to ssDNA also a β-hairpin is involved. In this connection it deserves to be mentioned that de Jong *et al.* (1989b) have proposed that in ssDNA binding residues 46 to 60, instead of residues 95 to 108 (*vide supra*) are involved. For two reasons we believe, however, that this assumption is incorrect. Firstly, the sequence proposed was not selected by the computer search programmes. Secondly, the sequence does not overlap with the amino acid stretch in which the DNA binding domain of g32p has been mapped.

E. coli RecA protein

Studies aimed to map the ssDNA binding domain(s) of RecA protein have made use of a number of different approaches, the results of which are not, however, very consistent. Deletion analysis have fairly strongly indicated that (a part of) this domain is located between residues 200 and 300 (Rusche *et al.*, 1985; Benedict and Kowalczykowski, 1988; Volodin, 1990). Also nucleotide sequence comparisons (Prasad and Chiu, 1987) have pointed into the direction that the RecA region running from residue 225 to 312 might contain (at least part of) the DNA binding domain. It is also within this region that we, but also others (de Jong *et al.*, 1989b), have found by computer analysis an amino acid sequence (residues 243-257) that is homologous to the DNA binding wing of gVp of Ff and IKe. Analyses of the recently published crystal structure of RecA protein have indicated that the two protein domains, consisting of residues 157-164 and 195-209, located on the surface of the protein, are highly conserved among bacterial RecA proteins (Roca and Cox, 1990; Story *et al.*, 1992). This plus the fact that missense mutations in each of these two domains affect DNA binding, have led Story *et al.* (1992) to propose that these domains are intimately involved in DNA binding and that the DNA is situated in the centre of the helical RecA•DNA filament. According to the crystal structure, the region we have found to be homologous to the DNA binding wing of gVp is situated in the core of the protein and consequently insufficiently exposed to be implicated in ssDNA binding. Since the observed homology between the RecA amino acid sequence (residues 243-257) and the DNA binding wing of Ff and IKe is of low statistical significance the alignment data are not in conflict with the latter conclusion. Despite the fact that the selected sequence is not involved in DNA binding it is, however, interesting to note that, like the DNA binding wing in gVp, it adopts a β-loop conformation in the native RecA protein (Story *et al.*, 1992).

I2-2 gVp and If1 DBP

The only biochemical information currently available about the ssDBPs encoded by the phages I2-2 and If1 is their amino acid sequence. We have been able to allocate in both proteins a stretch of amino acids that is homologous to the ssDNA binding wings of gVp of IKe and Ff.The statistical analysis of the alignment between residues 26-38 of the ssDBP of If1 and 16-28 of gVp of Ff shows that it is significant (z=3.1). Likewise, our analysis indicates that the amino acid sequence running from residue 79 to 92 in gVp of I2-2 is homologous to the amino acid sequences comprising the DNA binding wings in gVp of Ff and IKe (z=4.3 and z=4.4, respectively). Also between the indicated sequence in gVp of I2-2 and the Pf3 ssDBP sequence that runs from residue 12 to 24 a statistically significant homology was observed (z=3.5). Although there is no experimental evidence which supports the existence

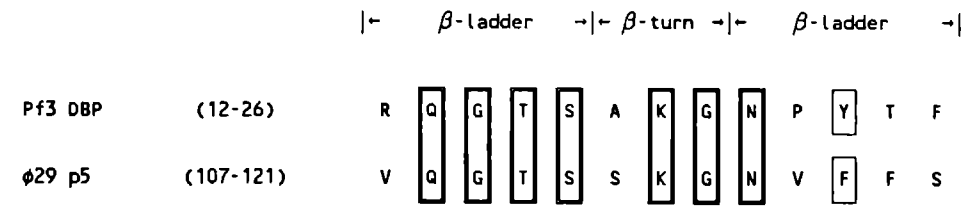


Figure 6.3 Proposed alignment of the amino acid sequences comprising the putative DNA binding loops of φ29 p5 protein and Pf3 DBP.
The similarities between the two sequences are indicated by boxes. Identical residues are boxed by double lines, whereas conservative replacements are boxed by single lines.
The amino acids are represented by one-letter symbols.

of a DNA binding wing in gVp of I2-2, the high *z*-values make us, however, feel confident to conclude that this must be the case. It remains remarkable however that, as also has been observed for the putative DNA binding wing in protein p5 of phage ϕ29 (*vide supra*), the DNA binding wing of gVp of I2-2 is situated in the C-terminal half and not in the N-terminal half of the protein, as is in gVp of Ff and IKe.

Vimentin

The intermediate filament protein vimentin is a prominent component of the cytoskeleton and nuclear envelope. The protein has been shown to bind *in vitro* with high affinity to ssDNA and RNA (Shoeman *et al.*, 1988). This interaction seems to be dictated primarily by the N-terminal domain of the protein, and both electrostatic and hydrophobic interactions are involved. Like gVp of Ff, binding of vimentin to ssDNA prevents nitration of the tyrosine residues in the protein (Traub *et al.*, 1992). Recently, Traub *et al.* (1992) have proposed that the N-terminal half of vimentin contains a ssDNA binding domain that is structurally and functionally similar to the DNA binding wing of gVp of Ff and IKe. Because of this very interesting hypothesis we decided to include the amino acid sequences of human/rodent and chicken vimentin proteins in our computer assisted analyses. These proteins exhibit strong identity at the level of amino acid sequences (human *vs.* rodent 97%, and human *vs.* chicken 87%).

In the N-terminal half of both human and rodent (data not shown) vimentin, amino acid sequences were found that are significantly homologous to the sequence of the DNA binding wing in gVp of Ff and IKe, suggesting indeed that these sequences play a role in ssDNA binding. However, chicken vimentin does not contain a sequence that is statistically significantly homologous to the DNA binding wing. The latter observation is extremely surprising particularly because it is to be expected that in all eukaryotic cells the evolutionarily related vimentins fulfil identical functions. Up to now we can only speculate about the meaning of these observations. Either they mean that the homologous sequences found in human and rodent vimentins are not involved in ssDNA binding at all, or that in ssDNA binding of chicken vimentin other amino acid stretches are involved. It furthermore might be possible that the observed *in vitro* ssDNA binding relies on an artefact and consequently that *in vivo* vimentin is not involved in ssDNA binding at all. Which of these hypothesis indeed is correct thus awaits more profound investigations.

FINAL DISCUSSION

Our analyses demonstrate that only the prokaryotic ssDBPs encoded by the phages I2-2, If1, Pf1, Pf3, T4, and ϕ29 and the eukaryotic intermediate filament protein vimentin, contain amino acid stretches with a statistically significant homology to the amino acid sequence of the DNA binding wing of gVp of Ff and IKe. These observations therefore strongly suggest that the DNA binding wing is indeed a recurrent DNA binding motif. The fact that the said motif has only been found in a fairly limited number of proteins is, nevertheless, very surprising. However, the fact that homologous amino acid stretches have been found in ssDBPs as distant as human vimentin and gVp of filamentous bacteriophages forms probably an indication that it finally might turn out that the DNA binding wing is in fact a much more universal occurring motif than at first glance is apparent from our computer assisted sequence analyses. In this context it furthermore deserves to be mentioned that despite the absence of significant amino acid sequence homology (overall identity 20%), structurally identical helix-turn-helix DNA binding motifs have been found in DBPs as distant as the phage lambda repressor and the POU-domain protein Oct-1 (Assa-Munt *et al.*, 1993; Dekker *et al.*, 1993). On the basis of these observations it might therefore well be that the DNA binding wing as present in gVp is a much more universal occurring ssDNA binding motif, but that its presence in most of the proteins analysed escaped detection because of degeneracy of the protein folding code.

REFERENCES

ALBERTS, B.M., AMODIO, F.J., JENKINS, M., GUTMANN, E.D., and FERRIS, F.L. (1968). Studies with DNA-cellulose chromatography. I. DNA-binding proteins from *Escherichia coli*. *Cold Spring Harbor Symp. Quant. Biol.* **33**: 289-305.

ALBERTS, B.M., and FREY, L. (1970). T4 bacteriophage gene 32: a structural protein in the replication and recombination of DNA. *Nature* **227**: 1313-1318.

ALBERTS, B., and HERRICK, G. (1971). DNA-cellulose chromatography. *Methods Enzym.* **21**: 198-217.

ALBERTS, B., FREY, L., and DELIUS, H. (1972). Isolation and characterization of gene 5 protein of filamentous bacterial viruses. *J. Mol. Biol.* **68**: 139-152.

ALMA, N.C.M., HARMSEN, B.J.M., HULL, W.E., VAN DER MAREL, G., VAN BOOM, J.H., and HILBERS, C.W. (1981a). Double-resonance experiments at 500 MHz on gene-5 protein and its complex with octadeoxyriboadenylic acid. *Biochemistry* **20**: 4419-4428.

ALMA, N.C.M., HARMSEN, B.J.M., HILBERS, C.W., VAN DER MAREL, G., and VAN BOOM, J.H. (1981b). 500 MHz ^{1}H NMR study of the role of lysines and arginines in the binding of gene-5 protein to oligoadenylic acids. *FEBS Lett.* **135**: 15-20.

ALMA, N.C.M., HARMSEN, B.J.M., VAN BOOM, J.H., VAN DER MAREL, G., and HILBERS, C.W. (1982). ^{1}H NMR studies of the binding of bacteriophage-M13-encoded gene-5 protein to oligo(deoxyadenylic acid)s of varying length. *Eur. J. Biochem.* **122**: 319-326.

ALMA, N.C.M., HARMSEN, B.J.M., VAN BOOM, J.H., VAN DER MAREL, G., and HILBERS, C.W. (1983a). A 500-MHz proton nuclear magnetic resonance study of the structure and structural alterations of gene-5 protein-oligo(deoxyadenylic acid) complexes. *Biochemistry* **22**: 2104-2115.

ALMA, N.C.M., HARMSEN, B.J.M., DE JONG, E.A.M., VAN DE VEN, J., and HILBERS, C.W. (1983b). Fluorescence studies of the complex formation between the gene 5 protein of bacteriophage M13 and polynucleotides. *J. Mol. Biol.* **163**: 47-62.

ANDERS, D.G. (1990). Nucleotide sequence of a cytomegalovirus single-stranded DNA-binding protein gene: comparison with alpha- and gamma-herpesvirus counterparts reveals conserved segments. *J. Gen. Virol.* **71**: 2451-2456.

ANDERSON, R.A., NAKASHIMA, Y., and COLEMAN, J.E. (1975). Chemical modifications of functional residues of fd gene 5 DNA-binding protein. *Biochemistry* **14**: 907-917.

ASSA-MUNT, N., MORTISHIRE-SMITH, R.J., AURORA, R., HERR, W., and WRIGHT, P.E. (1993). The solution structure of the Oct-1 POU-specific domain reveals a striking similarity to the bacteriophage λ repressor DNA-binding domain. *Cell* **73**: 193-205.

BAER, R., BANKIER, A.T., BIGGIN, M.D., DEININGER, P.L., FARRELL, P.J., GIBSON, T.J., HATFULL, G., HUDSON, G.S., SATCHWELL, S.C., SÉGUIN, C., TUFFNELL, P.S., and BARRELL, B.G. (1984). DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* **310**: 207-211.

BAX, A., and DAVIS, D.G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Res.* **65**: 355-360.

BAYER, M.E. (1968). Areas of adhesion between wall and membrane of *Escherichia coli*. *J. Gen. Microbiol.* **53**: 395-404.

BECK, E., SOMMER, R., AUERSWALD, E.A., KURZ, C., ZINK, B., OSTERBURG, G., and SCHALLER, H. (1978). Nucleotide sequence of bacteriophage fd DNA. *Nucleic Acids Res.* **5**: 4495-4503.

BECK, E., and ZINK, B. (1981). Nucleotide sequence and genome organisation of filamentous bacteriophages f1 and fd. *Gene* **16**: 35-58.

BECK, P.J., GONZALES, S., WARD, C.L., and MOLINEUX, I.J. (1989). Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9. *J. Mol. Biol.* **210**: 687-701.

BENEDICT, R.C., and KOWALCZYKOWSKI, S.C. (1988). Increase of the DNA strand assimilation activity of recA protein by removal of the C terminus and structure-function studies of the resulting protein fragment. *J. Biol. Chem.* **263**: 15513-15520.

BIAMONTI, G., BUVOLI, M., BASSI, M.T., MORANDI, C., COBIANCHI, F., and RIVA, S. (1989). Isolation of an active gene encoding human hnRNP protein A1. Evidence for alternative splicing. *J. Mol. Biol.* **207**: 491-503.

BIRNBOIM, H.C., and DOLY, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**: 1513-1523.

BLUMER, K.J., IVEY, M.R., and STEEGE, D.A. (1987). Translational control of phage f1 gene expression by differential activities of the gene V, VII, IX and VIII initiation sites. *J. Mol. Biol.* **197**: 439-451.

BOBST, A.M., LANGEMEIER, P.W., WARWICK-KOOCHAKI, P.E., BOBST, E.V., and IRELAND, J.C. (1982). Nucleic acid binding affinity of bacteriophage T4 gene 32 protein in the cooperative binding mode. *J. Biol. Chem.* **257**: 6184-6193.

BOBST, A.M., IRELAND, J.C., and BOBST, E.V. (1984). Nucleic acid binding affinity of fd gene 5 protein in the cooperative binding mode. *J. Biol. Chem.* **259**: 2130-2134.

BRADFORD, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**: 248-254.

BRADLEY, D.E. (1964). The structure of some bacteriophages associated with male strains of *Escherichia*

coli. J. Gen. Microbiol. **35**: 471-482.

BRADLEY, D.E., COETZEE, J.N., BOTHMA, T., and HEDGES, R.W. (1981). Phage X: a plasmid-dependent, broad host range, filamentous bacterial virus. *J. Gen. Microbiol.* **126**: 389-396.

BRADLEY, D.E., SIRGEL, F.A., COETZEE, J.N., HEDGES, R.W., and COETZEE, W.F. (1982). Phages C-2 and J: IncC and IncI plasmid-dependent phages, respectively. *J. Gen. Microbiol.* **128**: 2485-2498.

BRADLEY, D.E., COETZEE, J.N., and HEDGES, R.W. (1983). $IncI_2$ plasmids specify sensitivity to filamentous bacteriophage IKe. *J. Bacteriol.* **154**: 505-507.

BRAYER, G.D., and MCPHERSON, A. (1983). Refined structure of the gene 5 DNA binding protein from bacteriophage fd. *J. Mol. Biol.* **169**: 565-596.

BRAYER, G.D., and MCPHERSON, A. (1984a). Mechanism of DNA binding of the gene 5 protein of bacteriophage fd. *Biochemistry* **23**: 340-349.

BRAYER, G.D., and MCPHERSON, A. (1984b). Cooperative interactions of the gene 5 protein. *J. Biomol. Struct. Dynam.* **2**: 495-510.

BRAYER, G.D., and MCPHERSON, A. (1985). A model for intracellular complexation between gene-5 protein and bacteriophage fd DNA. *Eur. J. Biochem.* **150**: 287-296.

BRUTLAG, D., SCHEKMAN, R., and KORNBERG, A. (1971). A possible role for RNA polymerase in the initiation of M13 DNA synthesis. *Proc. Natl. Acad. Sci. USA* **68**: 2826-2829.

BULSINK, H., HARMSEN, B.J.M., and HILBERS, C.W. (1985) Specificity of the binding of bacteriophage M13 encoded gene-5 protein to DNA and RNA studied by means of fluorescence titrations. *J. Biomol. Struct. Dynam.* **3**: 227-247.

BULSINK, H., WIJNAENDTS VAN RESANDT, R.W., HARMSEN, B.J.M., and HILBERS, C.W. (1986). Different DNA-binding modes and cooperativities for bacteriophage M13 gene-5 protein revealed by means of fluorescence depolarisation studies. *Eur. J. Biochem.* **157**: 329-334.

BULSINK, H., HARMSEN, B.J.M., and HILBERS, C.W. (1988). DNA-binding properties of gene-5 protein encoded by bacteriophage M13. 2. Further characterization of the different binding modes for poly- and oligodeoxynucleic acids. *Eur. J. Biochem.* **176**: 597-608.

BUVOLI, M., BIAMONTI, G., TSOULFAS, P., BASSI, M.T., GHETTI, A., RIVA, S., and MORANDI, C. (1988). cDNA cloning of human hnRNP protein A1 reveals the existence of multiple mRNA isoforms. *Nucleic Acids Res.* **16**: 3751-3770.

CARNE, A., HILL, D.F., STOCKWELL, P.A., HUGHES, G., and PETERSEN, G.B. (1991). The putative single-stranded DNA-binding protein of the filamentous bacteriophage, If1. Amino acid sequence of the protein and structure of the gene. *Proc. R. Soc. Lond. B* **245**: 23-30.

CARO, L.G., and SCHNÖS, M. (1966). The attachment of the male-specific bacteriophage f1 to sensitive strains of *Escherichia coli. Proc. Natl. Acad. Sci. USA* **56**: 126-132.

CARPENTER, M.L., and KNEALE, G.G. (1991). Circular dichroism and fluorescence analysis of the interaction of Pf1 gene 5 protein with poly(dT). *J. Mol. Biol.* **217**: 681-689.

CASADABAN, M.J., and COHEN, S.N. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli. J. Mol. Biol.* **138**: 179-207.

CASAS-FINET, J.R., TOULMÉ, J.-J., SANTUS, R., and MAKI, A.H. (1988). A maximum of two tryptophan reidues in gene-32 protein from phage T4 undergo stacking interactions with single-stranded polynucleotides. *Eur. J. Biochem.* **172**: 641-646.

CASHMAN, J.S., and WEBSTER, R.E. (1979). Bacteriophage f1 infection of *Escherichia coli*: identification and possible processing of f1-specific mRNAs *in vivo. Proc. Natl. Acad. Sci. USA* **76**: 1169-1173.

CASHMAN, J.S., WEBSTER, R.E., and STEEGE, D.A. (1980). Transcription of bacteriophage f1. The major *in vivo* RNAs. *J. Biol. Chem.* **255**: 2554-2562.

CASSANI, G.R., and BOLLUM, F.J. (1969). Oligodeoxythymidylate:polydeoxyadenylate and oligodeoxyadenylate: polydeoxythymidylate interactions. *Biochemistry* **8**: 3928-3936.

CAVALIERI, S.J., NEET, K.E., and GOLDTHWAIT, D.A. (1976). Gene 5 protein of bacteriophage fd: a dimer which interacts co-operatively with DNA. *J. Mol. Biol.* **102**: 697-711.

CHANG, A.C.Y., and COHEN, S.N. (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* **134**: 1141-1156.

CHANG, L.M.S., LURIE, K., and PLEVANI, P. (1979). A stimulatory factor for yeast DNA polymerase. *Cold Spring Harbor Symp. Quant. Biol.* **43**: 587-595.

CHASE, J.W., MERRILL, B.M., and WILLIAMS, K.R. (1983). F sex factor encodes a single-stranded DNA binding protein (SSB) with extensive sequence homology to *Escherichia coli* SSB. *Proc. Natl. Acad. Sci. USA* **80**: 5480-5484.

CHASE, J.W., and WILLIAMS, K.R. (1986). Single-stranded DNA binding proteins required for DNA replication. *Ann. Rev. Biochem.* **55**: 103-136.

CHOU, P.Y., and FASMAN, G.D. (1978). Empirical predictions of protein conformation. *Ann. Rev. Biochem.* **47**: 251-276.

CHURCHILL, M.E.A., and TRAVERS, A.A. (1991). Protein motifs that recognize structural features of DNA. *TIBS* **16**: 92-97.

CLAESEN, C.A.A., DAEMEN, C.J.M., and TESSER, G.I. (1986). Synthesis of spin-labelled di- and tri-2-deoxyadenylates. *Recl. Trav. Chim. Pays-Bas* **105**: 116-123.

CLARK, A.J., and MARGULIES, A.D. (1965). Isolation and characterization of recombination-deficient mutants of *Escherichia coli* K12. *Proc. Natl. Acad. Sci. USA* **53**: 451-459.

CLEARY, J.M., and RAY, D.S. (1980). Replication of the plasmid pBR322 under the control of a cloned replication origin from the single-stranded DNA phage M13. *Proc. Natl. Acad. Sci. USA* **77**: 4638-4642.

CLEARY, J.M., and RAY, D.S. (1981). Deletion analysis

of the cloned replication origin region from bacteriophage M13. *J. Virol.* **40**: 197-203.

COBIANCHI, F., SENGUPTA, D.N., ZMUDZKA, B.Z., and WILSON, S.H., (1986). Structure of rodent helix-destabilizing protein revealed by cDNA cloning. *J. Biol. Chem.* **261**: 3536-3543.

COETZEE, J.N., BRADLEY, D.E., and HEDGES, R.W. (1982). Phages Iα and I2-2: IncI plasmid-dependent bacteriophages. *J. Gen. Microbiol.* **128**: 2797-2804.

COETZEE, J.N., BRADLEY, D.E., HEDGES, R.W., TWEEHUIZEN, M., and DU TOIT, L. (1987). Phage tf-1: a filamentous bacteriophage specific for bacteria harbouring the IncT plasmid pIN25. *J. Gen. Microbiol.* **133**: 953-960.

COETZEE, J.N., BRADLEY, D.E., DU TOIT, L., and HEDGES, R.W. (1988). Bacteriophage X-2: a filamentous phage lysing IncX-plasmid-harbouring bacterial strains. *J. Gen. Microbiol.* **134**: 2535-2541.

COLEMAN, J.E., ANDERSON, R.A., RATCLIFFE, R.G., and ARMITAGE, I.M. (1976). Structure of gene 5 protein-oligodeoxynucleotide complexes as determined by ^{1}H, ^{19}F, and ^{31}P nuclear magnetic resonance. *Biochemistry* **15**: 5419-5430.

COLEMAN, J.E., and ARMITAGE, I.M. (1978). Tyrosyl-base-phenylalanyl intercalation in gene 5 protein-DNA complexes: proton nuclear magnetic resonance of selectively deuterated gene 5 protein. *Biochemistry* **17**: 5038-5045.

COPELAND, T.D., MORGAN, M.A., and OROSZLAN, S. (1983). Complete amino acid sequence of the nucleic acid-binding protein of bovine leukemia virus. *FEBS Lett.* **156**: 37-40.

CUYPERS, T., VAN DER OUDERAA, F.J., and DE JONG, W.W. (1974). The amino acid sequence of gene 5 protein of bacteriophage M13. *Biochem. Biophys. Res. Commun.* **59**: 557-563.

DAI, H., CHIANG, K.-S, and KUO, T.-T (1980). Characterization of a new filamentous phage Cf from *Xanthomonas citri*. *J. Gen. Virol.* **46**: 277-289.

DAVISON, A.J., and SCOTT, J.E. (1986). The complete DNA sequence of varicella-zoster virus. *J. Gen. Virol.* **67**: 1759-1816.

DAY, L.A. (1973). Circular dichroism and ultraviolet absorption of a deoxyribonucleic acid binding protein of filamentous bacteriophage. *Biochemistry* **12**: 5329-5339.

DAY, L.A., and WISEMAN, R.L. (1978). A comparison of DNA packaging in the virions of fd, Xf, and Pf1. In: *The Single-Stranded DNA Phages* (D.T. Denhardt, D. Dressler, and D.S. Ray, eds.) pp. 605-625, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

DAY, L.A., MARZEC, C.J., REISBERG, S.A., and CASADEVALL, A. (1988). DNA packing in filamentous bacteriophages. *Ann. Rev. Biophys. Biophys. Chem.* **17**: 509-539.

DAYHOFF, M.O., SCHWARTZ, R.M., and ORCUTT, B.C. (1978). A model of evolutionary change in proteins. In: *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 3 (M.O. Dayhoff, ed.) pp. 345-352, National Biomedical Research Foundation, Washington D.C.

DAYHOFF, M.O., BARKER, W.C., and HUNT, L.T. (1983). Establishing homologies in protein sequences. *Methods Enzym.* **91**: 524-545.

DEKKER, N., COX, M., BOELENS, R., VERRIJZER, C.P., VAN DER VLIET, P.C., and KAPTEIN, R. (1993). Solution structure of the POU-specific DNA-binding domain of Oct-1. *Nature* **362**: 852-855.

DENHARDT, D.T., DRESSLER, D., and RAY, D.S. (eds.) (1978). *The Single-Stranded DNA Phages*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

DEVEREUX, J., HAEBERLI, P., and SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**: 387-395.

DOTTO, G.P., ENEA, V., and ZINDER, N.D. (1981). Functional analysis of bacteriophage f1 intergenic region. *Virology* **114**: 463-473.

DOTTO, G.P., HORIUCHI, K., and ZINDER, N.D. (1982). Initiation and termination of phage f1 plus-strand synthesis. *Proc. Natl. Acad. Sci. USA* **79**: 7122-7126.

DOTTO, G.P., and ZINDER, N.D. (1983). The morphogenetic signal of bacteriophage f1. *Virology* **130**: 252-256.

DOTTO, G.P., and ZINDER, N.D. (1984). Increased intracellular concentration of an initiator protein markedly reduces the minimal sequence required for initiation of DNA synthesis. *Proc. Natl. Acad. Sci. USA* **81**: 1336-1340.

DUNKER, A.K., and ANDERSON, E.A. (1975). The binding of the fd gene-5 protein to single-stranded nucleic acid. *Biochim. Biophys. Acta* **402**: 31-34.

DUNN, J.J., and STUDIER, F.W. (1981). Nucleotide sequence from the genetic left end of bacteriophage T7 DNA to the beginning of gene 4. *J. Mol. Biol.* **148**: 303-330.

VAN DUYNHOVEN, J.P.M., FOLKERS, P.J.M., STASSEN, A.P.M., HARMSEN, B.J.M., KONINGS, R.N.H., and HILBERS, C.W. (1990). Structure of the DNA binding wing of the gene-V encoded single-stranded DNA binding protein of the filamentous bacteriophage M13. *FEBS Lett.* **261**: 1-4.

VAN DUYNHOVEN, J.P.M., FOLKERS, P.J.M., PRINSE, C.W.J.M., HARMSEN, B.J.M., KONINGS, R.N.H., and HILBERS, C.W. (1992). Assignment of the ^{1}H NMR spectrum and secondary structure elucidation of the single-stranded DNA binding protein encoded by the filamentous bacteriophage IKe. *Biochemistry* **31**: 1254-1262.

VAN DUYNHOVEN, J.P.M., NOOREN, I.M.A., SWINKELS, D.W., FOLKERS, P.J.M., HARMSEN, B.J.M., KONINGS, R.N.H., TESSER, G.I., and HILBERS, C.W. (1993). Exploration of the single-stranded DNA-binding domains of the gene V proteins encoded by the filamentous bacteriophages IKe and M13 by means of spin-labeled oligonucleotide and lanthanide-chelate complexes. *Eur. J. Biochem.* **216**: 507-517.

VAN DYCK, E., FOURY, F., STILLMAN, B., and BRILL, S.J. (1992). A single-stranded DNA binding protein required for mitochondrial DNA replication in *S.*

cerevisiae is homologous to *E. coli* SSB. *EMBO J.* **11**: 3421-3430.

EDENS, L., KONINGS, R.N.H., and SCHOENMAKERS, J.G.G. (1975). Physical mapping of the central terminator for transcription on the bacteriophage M13 genome. *Nucleic Acids Res.* **2**: 1811-1820.

EDENS, L., KONINGS, R.N.H., and SCHOENMAKERS, J.G.G. (1978). A cascade mechanism of transcription in bacteriophage M13 DNA. *Virology* **86**: 354-367.

ENEA, V., and ZINDER, N.D. (1982). Interference resistant mutants of phage f1. *Virology* **122**: 222-226.

ERDILE, L.F., HEYER, W.-D., KOLODNER, R., and KELLY, T.J. (1991). Characterization of a cDNA encoding the 70-kDa single-stranded DNA-binding subunit of human replication protein A and the role of the protein in DNA replication. *J. Biol. Chem.* **266**: 12090-12098.

FANE, B., and KING, J. (1991). Intragenic suppressors of folding defects in the P22 tailspike protein. *Genetics* **127**: 263-277.

FERRARI, S., BATTINI, R., KACZMAREK, L., RITTLING, S., CALABRETTA, B., DE RIEL, J.K., PHILIPONIS, V., WEI, J.-F., and BASERGA, R. (1986). Coding sequence and growth regulation of the human vimentin gene. *Mol. Cell. Biol.* **6**: 3614-3620.

FIDANIÁN, H.M., and RAY, D.S. (1972). Replication of bacteriophage M13. VII. Requirement of the gene 2 protein for the accumulation of a specific RFII species. *J. Mol. Biol.* **72**: 51-63.

FOLKERS, P.J.M., VAN DUYNHOVEN, J.P.M., JONKER, A.J., HARMSEN, B.J.M., KONINGS, R.N.H., and HILBERS, C.W. (1991). Sequence-specific ^{1}H-NMR assignment and secondary structure of the Tyr41→His mutant of the single-stranded DNA binding protein, gene V protein, encoded by the filamentous bacteriophage M13. *Eur. J. Biochem.* **202**: 349-360.

FOLKERS, P.J.M., FOLMER, R.H.A., KONINGS, R.N.H., and HILBERS, C.W. (1993a). Overcoming the ambiguity problem encountered in the analysis of nuclear overhauser magnetic resonance spectra of symmetric dimer proteins. *J. Am. Chem. Soc.* **115**, 3798-3799.

FOLKERS, P.J.M., VAN DUYNHOVEN, J.P.M., VAN LIESHOUT, H.T.M., HARMSEN, B.J.M., VAN BOOM, J.H., TESSER, G.I., KONINGS, R.N.H., and HILBERS, C.W. (1993b). Exploring the DNA binding domain of gene V protein encoded by bacteriophage M13 with the aid of spin-labeled oligonucleotides in combination with ^{1}H-NMR. *Biochemistry* **32**: 9407-9416.

FOLKERS, P.J.M., NILGES, M., FOLMER, R.H.A., KONINGS, R.N.H., and HILBERS, C.W. (1994). The solution structure of the Tyr41→His mutant of the single-stranded DNA binding protein encoded by gene V of the filamentous bacteriophage M13. *J. Mol. Biol.* **236**: 229-246.

FREEMONT, P.S., LANE, A.N., and SANDERSON, M.R. (1991). Structural aspects of protein-DNA recognition. *Biochem. J.* **278**: 1-23.

FULFORD, W., and MODEL, P. (1984). Specificity of translational regulation by two DNA-binding proteins. *J. Mol. Biol.* **173**: 229-226.

FULFORD, W., and MODEL, P. (1988a). Bacteriophage f1 DNA replication genes. II. The roles of gene V protein and gene II protein in complementary strand synthesis. *J. Mol. Biol.* **203**: 39-48.

FULFORD, W., and MODEL, P. (1988b). Regulation of bacteriophage f1 DNA replication. I. New functions for genes II and X. *J. Mol. Biol.* **203**: 49-62.

GARSSEN, G.J., HILBERS, C.W., SCHOENMAKERS, J.G.G., and VAN BOOM, J.H. (1977). Studies on DNA unwinding. Proton and phosphorus nuclear-magnetic-resonance studies of gene V protein from bacteriophage M13, interacting with d(pC-G-C-G). *Eur. J. Biochem.* **81**: 453-463.

GARSSEN, G.J., KAPTEIN, R., SCHOENMAKERS, J.G.G., and HILBERS, C.W. (1978). A photo-CIDNP study of the interaction of oligonucleotides with gene-5 protein of bacteriophage M13. *Proc. Natl. Acad. Sci. USA* **75**: 5281-5285.

GARSSEN, G.J., TESSER, G.I., SCHOENMAKERS, J.G.G., and HILBERS, C.W. (1980). NMR studies of the interaction of gene-V protein of bacteriophage M13 with oligonucleotides. *Biochem. Biophys. Acta.* **607**: 361-371.

GEIDER, K., and KORNBERG, A. (1974). Conversion of the M13 viral single strand to the double-stranded replicative forms by purified proteins. *J. Biol. Chem.* **249**: 3999-4005.

GEIDER, K. (1978). Interaction of DNA with DNA-binding proteins. Protein exchange and complex stability. *Eur. J. Biochem.* **87**: 617-622.

GEIDER, K., BECK, E., and SCHALLER, H. (1978). An RNA transcribed from DNA at the origin of phage fd single strand to replicative form conversion. *Proc. Natl. Acad. Sci. USA* **75**: 645-649.

GHRIR, R., LECAER, J.-P, DUFRESNE, C., and GUERIDE, M. (1991). Primary structure of the two variants of *Xenopus laevis* mtSSB, a mitochondrial DNA binding protein. *Arch. Biochem. Biophys.* **291**: 395-400.

GILBERT, W., and DRESSLER, D. (1968). DNA replication: the rolling circle model. *Cold Spring Harbor Symp. Quant. Biol.* **33**: 473-484.

GOLDSMITH, M.E., and KONIGSBERG, W.H. (1977). Absorption protein of the bacteriophage fd: isolation, molecular properties, and location in the virus. *Biochemistry* **16**: 2686-2694.

GRANT, R.A., LIN, T.-C., KONIGSBERG, W., and WEBSTER, R.E. (1981). Structure of the filamentous bacteriophage f1. Location of the A, C, and D minor coat proteins. *J. Biol. Chem.* **256**: 539-546.

GRANT, R.B., WHITELEY, M.H., and SHAPLEY, A.J. (1978). Plasmids of incompatibility group P code for the capacity to propagate bacteriophage IKe. *J. Bacteriol.* **136**: 808-811.

GRAY, C.P., SOMMER, R., POLKE, C., BECK, E., and SCHALLER, H. (1978). Structure of the origin of DNA replication of bacteriophage fd. *Proc. Natl. Acad. Sci. USA* **75**: 50-53.

GRAY, C.W., KNEALE, G.G., LEONARD, K.R., SIEGRIST, H., and MARVIN, D.A. (1982a). A nucleoprotein complex in bacteria infected with Pf1 filamentous virus: identification and electron microscopic

analysis. *Virology* **116**: 40-52.

GRAY, C.W. (1989). Three-dimensional structure of complexes of single-stranded DNA-binding proteins with DNA. IKe and fd gene 5 proteins form left-handed helices with single-stranded DNA. *J. Mol. Biol.* **208**: 57-64.

GRAY, D.M., GRAY, C.W., and CARLSON, R.D. (1982b). Neutron scattering data on reconstituted complexes of fd deoxyribonucleic acid and gene 5 protein show that the deoxyribonucleic acid is near the center. *Biochemistry* **21**: 2702-2713.

GRAY, D.M., SHERRY, A.D., TEHERANI, J., and KANSY, J.W. (1984). Reductive methylation of the lysyl residues in the fd gene 5 DNA-binding protein: CD and ^{13}C NMR results on the modified protein. *J. Biomol. Struct. Dynam.* **2**: 77-91.

GREULICH, K.O., WIJNAENDTS VAN RESANDT, R.W., and KNEALE, G.G. (1985). Time resolved fluorescence of bacteriophage Pf1 DNA binding protein and its complex with DNA. *Eur. Biophys. J.* **11**: 195-201.

GRIBSKOV, M., and BURGESS, R.R. (1986). Sigma factors from *E. coli*, *B. subtilis*, phage SPO1, and phage T4 are homologous proteins. *Nucleic Acids Res.* **14**: 6745-6763.

GRIESINGER, C., OTTING, G., WÜTHRICH, K., and ERNST, R.R. (1988). Clean TOCSY for ^{1}H spin system identification in macromolecules. *J. Am. Chem. Soc.* **110**: 7870-7872.

HARLOW, E., and LANE, D. (1988). *Antibodies: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

HARRISON, S.C. (1991). A structural taxonomy of DNA-binding domains. *Nature* **353**: 715-719.

HAYNES, S.R., REBBERT, M.L., MOZER, B.A., FORQUIGNON, F., and DAWID, I.B. (1987). *pen* repeat sequences are GGN clusters and encode a glycine-rich domain in a *Drosophila* cDNA homologous to the rat helix destabilizing protein. *Proc. Natl. Acad. Sci. USA* **84**: 1819-1823.

HENDERSON, L.E., COPELAND, T.D., SOWDER, R.C., SMYTHERS, G.W., and OROSZLAN, S. (1981). Primary structure of the low molecular weight nucleic acid-binding proteins of murine leukemia viruses. *J. Biol. Chem.* **256**: 8400-8406.

HEYER, W.-D., RAO, M.R.S., ERDILE, L.F., KELLY, T.J., and KOLODNER, R.D. (1990). An essential *Saccharomyces cerevisiae* single-stranded DNA binding protein is homologous to the large subunit of human RP-A. *EMBO J.* **9**: 2321-2329.

HILL, D.F., and PETERSEN, G.B. (1982). Nucleotide sequence of bacteriophage f1 DNA. *J. Virol.* **44**: 32-46.

HILL, D.F., SHORT, N.J., PERHAM, R.N., and PETERSEN, G.B. (1991). DNA sequence of the filamentous bacteriophage Pf1. *J. Mol. Biol.* **218**: 349-364.

HOFSCHNEIDER, P.H. (1963). Untersuchungen über "kleine" *E. coli* K12 Bakteriophagen. *Z. Naturforschung* **18b**: 203-210.

HOKE, G.D., PAVCO, P.A., LEDWITH, B.J., and VAN TUYLE, G.C. (1990). Structural and functional studies of the rat mitochondrial single strand DNA binding protein P16. *Arch. Biochem. Biophys.* **282**: 116-124.

HOLWITT, E., and KRASNA, A.I. (1982). Interaction of gene 5 protein with DNA. *Arch. Biochem. Biophys.* **214**: 792-805.

VAN DEN HONDEL, C.A., PENNINGS, L., and SCHOENMAKERS, J.G.G. (1976). Restriction-enzyme-cleavage maps of bacteriophage M13. Existence of an intergenic region on the M13 genome. *Eur. J. Biochem.* **68**: 55-70.

HORII, T., OGAWA, T., and OGAWA, H. (1980). Organization of the *recA* gene of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **77**: 313-317.

HORIUCHI, K., and ZINDER, N.D. (1976). Origin and direction of synthesis of bacteriophage f1 DNA. *Proc. Natl. Acad. Sci. USA* **73**: 2341-2345.

HORIUCHI, K., RAVETCH, J.V., and ZINDER, N.D. (1979). DNA replication of bacteriophage f1 *in vivo*. *Cold Spring Harbor Symp. Quant. Biol.* **43**: 389-399.

HORWITZ, A.H., HEFFERNAN, L., MORANDI, C., LEE, J.-H., TIMKO, J., and WILCOX, G. (1981). DNA sequence of the *araBAD-araC* controlling region in *Salmonelle typhimurium* LT2. *Gene* **14**: 309-319.

HOWLAND, C.J., REES, C.E.D., BARTH, P.T., and WILKINS, B.M. (1989). The *ssb* gene of plasmid ColIb-P9. *J. Bacteriol.* **171**: 2466-2473.

HUTCHINSON, D.L., BARNETT, B.L., and BOBST, A.M. (1990). Gene 5 protein-DNA complex: modeling binding interactions. *J. Biomol. Struct. Dynam.* **8**: 1-9.

IVEY-HOYLE, M., and STEEGE, D.A. (1989). Translation of phage f1 gene VII occurs from an inherently defective initiation site made functional by coupling. *J. Mol. Biol.* **208**: 233-244.

JOHNSTON, S., LEE, J.-H., and RAY, D.S. (1985). High-level expression of M13 gene II protein from an inducible polycistronic messenger RNA. *Gene* **34**: 137-145.

DE JONG, E.A.M., HARMSEN, B.J.M., KONINGS, R.N.H., and HILBERS, C.W. (1987a). Binding of IKe gene 5 protein to polynucleotides. Fluorescence binding experiments of IKe gene 5 protein and mutual cooperativity of IKe and M13 gene 5 proteins. *Biochemistry* **26**: 2039-2046.

DE JONG, E.A.M., KONINGS, R.N.H., HARMSEN, B.J.M., PRINSE, C.W.J.M., and HILBERS, C.W. (1987b). ^{1}H-NMR studies on the gene-5-encoded single-stranded DNA binding protein of the filamentous bacteriophage IKe. General spectral and structural features. *Eur. J. Biochem.* **167**: 563-572.

DE JONG, E.A.M., VAN DUYNHOVEN, J.P.M., HARMSEN, B.J.M., KONINGS, R.N.H., and HILBERS, C.W. (1989a). Two-dimensional ^{1}H nuclear magnetic resonance studies on the gene V-encoded single-stranded DNA-binding protein of the filamentous bacteriophage IKe. I. Structure elucidation of the DNA-binding wing. *J. Mol. Biol.* **206**: 119-132.

DE JONG, E.A.M., VAN DUYNHOVEN, J.P.M., HARMSEN B.J.M., TESSER, G.I., KONINGS, R.N.H., and HILBERS, C.W. (1989b). Two-dimensional ^{1}H nuclear magnetic resonance studies on the gene V-encoded single-stranded DNA-binding protein

of the filamentous bacteriophage IKe. II. Characterization of the DNA-binding wing with the aid of spin-labelled oligonucleotides. *J. Mol. Biol.* **206**: 133-152.

JUKES, T.H., and CANTOR, C.R. (1969). Evolution of protein molecules. In: *Mammalian Protein Metabolism* (H.N. Munro, ed.) pp. 21-132, Academic Press, New York.

KANSY, J.W., CLACK, B.A., and GRAY, D.M. (1986). The binding f fd gene 5 protein to polynucleotides: evidende from CD measurements for two binding modes. *J. Biomol. Struct. Dynam.* **3**: 1079-1110.

KAO, S.-C., BOBST, E.V., PAULY, G.T., and BOBST, A.M. (1985). Molecular dynamics in protein-single stranded DNA complexes. Two distinct nucleoside mobilities in poly(deoxythymidylic acid)-gene 5 protein complexes. *J. Biomol. Struct. Dynam.* **3**: 261-268.

KHATOON, H., IYER, R.V., and IYER, V.N. (1972). A new filamentous bacteriophage with sex-factor specificity. *Virology* **48**: 145-155.

KING, G.C., and COLEMAN, J.E. (1987). Two-dimensional ^1H NMR of gene 5 protein indicates that only two aromatic rings interact significantly with oligodeoxynucleotide bases. *Biochemistry* **26**: 2929-2937.

KING, G.C., and COLEMAN, J.E. (1988). The Ff gene 5 protein-d$(pA)_{40\text{-}60}$ complex: ^1H NMR supports a localized base-binding model. *Biochemistry* **27**: 6947-6953.

KITCHINGMAN, G.R. (1985). Sequence of the DNA-binding protein of a human subgroup E adenovirus (type 4): comparisons with subgroup A (type 12), subgroup B (type 7), and subgroup C (type 5). *Virology* **146**: 90-101.

KNEALE, G.G., FREEMAN, R., and MARVIN, D.A. (1982). Pf1 bacteriophage replication-assembly complex. X-ray fibre diffraction and scanning transmission electron microscopy. *J. Mol. Biol.* **156**: 279-292.

KNEALE, G.G. (1983). Dissociation of the Pf1 nucleoprotein assembly complex and characterisation of the DNA binding protein. *Biochim. Biophys. Acta* **739**: 216-224.

KNEALE, G.G., and WIJNAENDTS VAN RESANDT, R.W. (1985). Time-resolved fluorescence of bacteriophage Pf1 DNA-binding protein. Determination of oligonucleotide and polynucleotide binding parameters. *Eur. J. Biochem.* **149**: 85-93.

KOKOSKA, R.J., BLUMER, K.J., and STEEGE, D.A. (1990). Phage f1 mRNA processing in *Escherichia coli*: search for the upstream products of endonuclease cleavage, requirement for the product of the altered mRNA stability (*ams*) locus. *Biochimie* **72**: 803-811.

KONINGS, R.N.H., HULSEBOS, T., and VAN DEN HONDEL, C.A. (1975). Identification and characterization of the *in vitro* synthesized gene products of bacteriophage M13. *J. Virol.* **15**: 570-584.

KONINGS, R.N.H., LUITEN, R.G.M., and PEETERS, B.P.H. (1986). Mike, a chimeric filamentous phage designed for the separate production of either DNA strand of pKUN vector plasmids by F^+ cells. *Gene* **46**: 269-276.

KONINGS, R.N.H., VERHOEVEN, E.J.M., and PEETERS, B.P.H. (1987). pKUN, vectors for the separate production of both DNA strands of recombinant plasmids. *Methods Enzym.* **153**: 12-34.

KORNBERG, A., and BAKER, T.A. (1992). DNA-binding proteins. In: *DNA Replication* (A. Kornberg, and T.A. Baker, eds.) pp. 323-354, W.H. Freeman and Company, New York.

KRISCH, H.M., and ALLET, B. (1982). Nucleotide sequences involved in bacteriophage T4 gene 32 translational self-regulation. *Proc. Natl. Acad. Sci. USA* **79**: 4937-4941.

KRUIJER, W., VAN SCHAIK, F.M.A., and SUSSENBACH, J.S. (1981). Structure and organization of the gene coding for the DNA binding protein of adenovirus type 5. *Nucleic Acids Res.* **9**: 4439-4457.

KRUIJER, W., VAN SCHAIK, F.M.A., and SUSSENBACH, J.S. (1982). Nucleotide sequence of the gene encoding adenovirus type 2 DNA binding protein. u*Nucleic Acids Res.* **10**: 4493-4500.

KRUIJER, W., VAN SCHAIK, F.M.A., SPEIJER, J.G., and SUSSENBACH, J.S. (1983). Structure and function of adenovirus DNA binding protein: comparison of the amino acid sequences of the Ad5 and Ad12 proteins derived from the nucleotide sequence of the corresponding genes. *Virology* **128**: 140-153.

KUO, T.-T., HUANG, T.-C., and CHOW, T.-Y. (1969). A filamentous bacteriophage from *Xanthomonas oryzae*. *Virology* **39**: 548-555.

KUO, T.-T., LIN, Y.-H., HUANG, C.-M., CHANG, S.-F., DAI, H., and FENG, T.-Y. (1987). The lysogenic cycle of the filamentous phage Cf1t from *Xanthomonas campestris* pv. *citri*. *Virology* **156**: 305-312.

KUO, T.-T., TAN, M.-S., SU, M.-T., and YANG, M.-K. (1991). Complete nucleotide sequence of filamentous phage Cf1c from *Xanthomonas campestris* pv. *citri*. *Nucleic Acids Res.* **19**: 2498.

LAEMMLI, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

LAKOWICZ, J.R. (1983). *Principles of Fluorescence Spectroscopy*, Plenum Press, New York.

LIANG, H., and TERWILLIGER, T.C. (1991). Reversible denaturation of the gene V protein of bacteriophage f1. *Biochemistry* **30**: 2772-2782.

LIN, H.-C., LEI, S.-P., and WILCOX, G. (1985). The *araBAD* operon of *Salmonella typhimurium* LT2. I. Nucleotide sequence of *araB* and primary structure of its product, ribulokinase. *Gene* **34**: 111-122.

LIN, N.S.-C., and PRATT, D. (1972). Role of bacteriophage M13 gene 2 in viral DNA replication. *J. Mol. Biol.* **72**: 37-49.

LIN, N.S.-C., and PRATT, D. (1974). Bacteriophage M13 gene 2 protein: increasing its yield in infected cells, and identification and localization. *Virology* **61**: 334-342.

LIN, T.-C., WEBSTER, R.E., and KONIGSBERG, W. (1980). Isolation and characterization of the C and D proteins coded by gene IX and gene VI in the filamentous bacteriophage f1 and fd. *J. Biol. Chem.* **255**: 10331-10337.

LINDBERG, G., KOWALCZYKOWSKI, S.C., RIST, J.K.,

SUGINO, A., and ROTHMAN-DENES, L.B. (1989). Purification and characterization of the coliphage N4-coded single-stranded DNA binding protein. *J. Biol. Chem.* **264**: 12700-12708.

LIPMAN, D.J., and PEARSON, W.R. (1985). Rapid and sensitive protein similarity searches. *Science* **227**: 1435-1441.

LOEB, T. (1960). Isolation of a bacteriophage specific for the F^+ and Hfr mating types of *Escherichia coli* K-12. *Science* **131**, 932-933.

LOPEZ, J., and WEBSTER, R.E. (1985). Assembly site of bacteriophage f1 corresponds to adhesion zones between the inner and outer membranes of the host cell. *J. Bacteriol.* **163**: 1270-1274.

LUITEN, R.G.M., PUTTERMAN, D.G., SCHOENMAKERS, J.G.G., KONINGS, R.N.H., and DAY, L.A. (1985). Nucleotide sequence of the genome of Pf3, an IncP-1 plasmid-specific filamentous bacteriophage of *Pseudomonas aeruginosa*. *J. Virol.* **56**: 268-276.

MAEDA, K., KNEALE, G.G., TSUGITA, A., SHORT, N.J., PERHAM, R.N., HILL, D.F., and PETERSEN, G.B. (1982). The DNA-binding protein of Pf1 filamentous bacteriophage: amino-acid sequence and structure of the gene. *EMBO J.* **1**: 255-261.

MAHOUNGOU, C., GHRIR, R., LECAER, J.P., MIGNOTTE, B., and BARAT-GUERIDE, M. (1988). The amino-terminal sequence of the *Xenopus laevis* mitochondrial SSB is homologous to that of the *Escherichia coli* protein. *FEBS Lett.* **235**: 267-270.

MAKOWSKI, L. (1984). Structural diversity in filamentous bacteriophages. In: *Biological Macromolecules and Assemblies*, vol. 1, *Virus Structures* (F.A. Jurnak, and A. McPherson, eds.) pp 203-253, John Wiley and Sons, New York.

MARION, D., and WÜTHRICH, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^1H-^1H spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* **113**: 967-974.

MARSTON, F.A.O. (1986). The purification of eukaryotic polypeptides synthesized in *Escherichia coli*. *Biochem. J.* **240**: 1-12.

MARTÍN, G., and SALAS, M. (1988). Characterization and cloning of gene 5 of *Bacillus subtilis* phage ϕ29. *Gene* **67**: 193-201.

MARTÍN, G., LÁZARO, J.M., MÉNDEZ, E., and SALAS, M. (1989). Characterization of the phage ϕ29 protein p5 as a single-stranded DNA binding protein. Function in ϕ29 DNA-protein p3 replication. *Nucleic Acids Res.* **17**: 3663-3672.

MARVIN, D.A., and HOFFMANN-BERLING, H. (1963). Physical and chemical properties of two new small bacteriophages. *Nature* **197**: 517-518.

MARVIN, D.A., and HOHN, B. (1969). Filamentous bacterial viruses. *Bacteriol. Rev.* **33**: 172-209.

MARVIN, D.A., WISEMAN, R.L., and WACHTEL, E.J. (1974a). Filamentous bacterial viruses. XI. Molecular architecture of the class II (Pf1, Xf) virion. *J. Mol. Biol.* **82**: 121-138.

MARVIN, D.A., PIGRAM, W.J., WISEMAN, R.L., WACHTEL, E.J., and MARVIN, F.J. (1974b). Filamentous bacterial viruses. XII. Molecular architecture of the class I (fd, If1, IKe) virion. *J. Mol. Biol.* **88**: 581-600.

MARVIN, D.A., and FOLKHARD, W. (1986). Structure of F-pili: Reassessment of the symmetry. *J. Mol. Biol.* **191**: 299-300.

MAZUR, B.J., and MODEL, P. (1973). Regulation of coliphage f1 single-stranded DNA synthesis by a DNA-binding protein. *J. Mol. Biol.* **78**: 285-300.

MAZUR, B.J., and ZINDER, N.D. (1975). The role of gene V protein in f1 single-strand synthesis. *Virology* **68**: 490-502.

MCGHEE, J.D., and VON HIPPEL, P.H. (1974). Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.* **86**: 469-489.

MCPHEETERS, D.S., GOSCH, G., and GOLD, L. (1988). Nucleotide sequences of the bacteriophage T2 and T6 gene 32 mRNAs. *Nucleic Acids Res.* **16**: 9341.

MCPHERSON, A., MOLINEUX, I., and RICH, A. (1976). Crystallization of a DNA-unwinding protein: preliminary X-ray analysis of fd bacteriophage gene 5 product. *J. Mol. Biol.* **106**: 1077-1081.

MCPHERSON, A. (1976). Crystallization of proteins from polyethylene glycol. *J. Biol. Chem.* **251**: 6300-6303.

MERRILL, B.M., LOPRESTI, M.B., STONE, K.L., and WILLIAMS, K.R. (1987). Amino acid sequence of UP1, an hnRNP-derived single-stranded nucleic acid binding protein from calf thymus. *Int. J. Peptide Protein Res.* **29**: 21-39.

MEYER, R.R., and LAINE, P.S. (1990). The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiol. Rev.* **54**: 342-380.

MEYER, T.F., and GEIDER, K. (1978). Isolation and partial characterization of gene-II protein from fd-infected cells. In: *The Single-Stranded DNA Phages* (D.T. Denhardt, D. Dressler, and D.S. Ray, eds.) pp 389-392, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

MEYER, T.F., GEIDER, K., KURZ, C., and SCHALLER, H. (1979). Cleavage site of bacteriophage fd gene II-protein in the origin of viral strand replication. *Nature* **278**: 365-367.

MEYER, T.F., and GEIDER, K. (1979a). Bacteriophage fd gene II-protein. I. Purification, involvement in RF replication, and the expression of gene II. *J. Biol. Chem.* **254**: 12636-12641.

MEYER, T.F., and GEIDER, K. (1979b). Bacteriophage fd gene II-protein. II. Specific cleavage and relaxation of supercoiled RF from filamentous phages. *J. Biol. Chem.* **254**: 12642-12646.

MEYER, T.F., and GEIDER, K. (1981). Cloning of bacteriophage fd gene 2 and construction of a plasmid dependent on fd gene 2 protein. *Proc. Natl. Acad. Sci. USA* **78**: 5416-5420.

MEYNELL, G.G., and LAWN, A.M. (1968). Filamentous phages specific for the *I* sex factor. *Nature* **217**: 1184-1186.

MICHEL, B., and ZINDER, N.D. (1989a). Translational repression in bacteriophage f1: characterization of

the gene V protein target on the gene II mRNA. *Proc. Natl. Acad. Sci. USA* **86**: 4002-4006.

MICHEL, B., and ZINDER, N.D. (1989b). *In vitro* binding of the bacteriophage f1 gene V protein to the gene II RNA-operator and its DNA analog. *Nucleic Acids Res.* **17**, 7333-7344.

MILLER, J.H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

MINAMISHIMA, Y., TAKEYA, K., OHNISKI, Y., and AMAKO, K. (1968). Physicochemical and biological properties of fibrous *Pseudomonas* bacteriophages. *J. Virol.* **2**: 208-213.

MITRAKI, A., and KING, J. (1989). Protein folding intermediates and inclusion body formation. *Bio/Technology* **7**: 690-697.

MIYATA, T., and YASUNAGA, T. (1980). Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *J. Mol. Evol.* **16**: 23-36.

MODEL, P., MCGILL, C., MAZUR, B., and FULFORD, W.D. (1982). The replication of bacteriophage f1: gene V protein regulates the synthesis of gene II protein. *Cell* **29**: 329-335.

MODEL, P., and RUSSEL, M. (1988). Filamentous bacteriophage. In: *The Bacteriophages*, vol. 2 (R. Calendar, ed.) pp. 375-456, Plenum Publishing Corp, New York.

MOLINEUX, I.J., FRIEDMAN, S., and GEFTER, M.L. (1974). Purification and properties of the *Escherichia coli* deoxyribonucleic acid-unwinding protein. Effects on deoxyribonucleic acid synthesis *in vitro*. *J. Biol. Chem.* **249**: 6090-6098.

MORRIS, G.A., and FREEMAN, R. (1978). Selective excitation in Fourier transform nuclear magnetic resonance. *J. Magn. Res.* **29**: 433-462.

MOSES, P.B., and MODEL, P. (1984). A rho-dependent transcription termination signal in bacteriophage f1. *J. Mol. Biol.* **172**: 1-22.

MYERS, R.M., LERMAN, L.S., and MANIATIS, T. (1985). A general method for saturation mutagenesis of cloned DNA fragments. *Science* **229**: 242-247.

NAKASHIMA, Y, DUNKER, A.K., MARVIN, D.A., and KONIGSBERG, W. (1974). The amino acid sequence of a DNA binding protein, the gene 5 product of fd filamentous bacteriophage. *FEBS Lett.* **40**: 290-292, and **43**: 125.

NEEDLEMAN, S.B., and WUNSCH, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**: 443-453.

O'CONNOR, T.P., and COLEMAN, J.E. (1983). Proton nuclear magnetic resonance (500 MHz) of mono-, di-, tri-, and tetradeoxynucleotide complexes of gene 5 protein. *Biochemistry* **22**: 3375-3381.

OEY, J.L., and KNIPPERS, R. (1972). Properties of the isolated gene 5 protein of bacteriophage fd. *J. Mol. Biol.* **68**: 125-138.

PABO, C.O., and SAUER, R.T. (1992). Transcription factors: structural families and principles of DNA recognition. *Ann. Rev. Biochem.* **61**: 1053-1095.

PAN, T., KING, G.C., and COLEMAN, J.E (1989). Comparison of cooperative and isolated site binding of T4 gene 32 protein to ssDNA by ^{1}H NMR. *Biochemistry* **28**: 8833-8839.

PEARSON, W.R., and LIPMAN, D.J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**: 2444-2448.

PEARSON, W.R. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzym.* **183**: 63-98.

PEETERS, B.P.H., KONINGS, R.N.H., and SCHOENMAKERS, J.G.G. (1983). Characterization of the DNA binding protein encoded by the N-specific filamentous *Escherichia coli* phage IKe. Binding properties of the protein and nucleotide sequence of the gene. *J. Mol. Biol.* **169**: 197-215.

PEETERS, B.P.H., PETERS, R.M., SCHOENMAKERS, J.G.G., and KONINGS, R.N.H. (1985). Nucleotide sequence and genetic organization of the genome of the N-specific filamentous bacteriophage IKe. Comparison with the genome of the F-specific filamentous phages M13, fd and f1. *J. Mol. Biol.* **181**: 27-39.

PLYTE, S.E., and KNEALE, G.G. (1991). The role of tyrosine residues in the DNA-binding site of the Pf1 gene 5 protein. *Prot. Engineering* **4**: 553-560.

PÖRSCHKE, D., and RAUH, H. (1983). Cooperative, excluded-site binding and its dynamics for the interaction of gene 5 protein with polynucleotides. *Biochemistry* **22**: 4737-4745.

PRASAD, B.V.V., and CHIU, W. (1987). Sequence comparison of single-stranded DNA binding proteins and its structural implications. *J. Mol. Biol.* **193**: 579-584.

PRATT, D., and ERDAHL, W.S. (1968). Genetic control of bacteriophage M13 DNA synthesis. *J. Mol. Biol.* **37**: 181-200.

PRATT, D., LAWS, P., and GRIFFITH, J. (1974). Complex of bacteriophage M13 single-stranded DNA and gene 5 protein. *J. Mol. Biol.* **82**: 425-439.

PRETORIUS, H.T., KLEIN, M., and DAY, L.A. (1975). Gene V protein of fd bacteriophage. Dimer formation and the role of tyrosyl groups in DNA binding. *J. Biol. Chem.* **250**: 9262-9269.

PRIGODICH, R.V., CASAS-FINET, J., WILLIAMS, K.R., KONIGSBERG, W., and COLEMAN, J.E. (1984). ^{1}H NMR (500 MHz) of gene 32 protein-oligonucleotide complexes. *Biochemistry* **23**: 522-529.

PRIGODICH, R.V., SHAMOO, Y., WILLIAMS, K.R., CHASE, J.W., KONIGSBERG, W.H., and COLEMAN, J.E. (1986). ^{1}H NMR (500 MHz) identification of aromatic residues of gene 32 protein involved in DNA binding by use of protein containing perdeuterated aromatic residues and by site-directed mutagenesis. *Biochemistry* **25**: 3666-3672.

PUTTERMAN, D.G., CASADEVALL, A., BOYLE, P.D., YANG, H.-L., FRANGIONE, B., and DAY, L.A. (1984). Major coat protein and single-stranded DNA-binding protein of filamentous virus Pf3. *Proc. Natl. Acad. Sci. USA* **81**: 699-703.

QUINN, C.O., and KITCHINGMAN, G.R. (1984). Sequence of the DNA-binding protein gene of a human subgroup B adenovirus (type 7). Compar-

isons with subgroup C (type 5) and subgroup A (type 12). *J. Biol. Chem.* **259**: 5003-5009.

QUINN, J.P., and MCGEOCH, D.J. (1985). DNA sequence of the region in the genome of herpes simplex virus type 1 containing the genes for DNA polymerase and the major DNA binding protein. *Nucleic Acids Res.* **13**: 8143-8163.

RADDING, C.M. (1982). Homologous pairing and strand exchange in genetic recombination. *Ann. Rev. Genet.* **16**: 405-437.

RANCE, M., SØRENSEN, O.W., BODENHAUSEN, G., WAGNER, G., ERNST, R.R., and WÜTHRICH, K. (1983). Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.* **117**: 479-485.

RASCHED, I., and POHL, F.M. (1974). Oligonucleotides and the quaternary structure of gene-5 protein from filamentous bacteriophage. *FEBS Lett.* **46**: 115-118.

RAY, D.S. (1977). Replication of filamentous bacteriophages. *Comprehensive Virol.* **7**: 105-178.

RECORD, M.T., LOHMAN, T.M., and DE HASETH, P. (1976). Ion effects on ligand-nucleic acid interactions. *J. Mol. Biol.* **107**: 145-158.

RECORD, M.T., ANDERSON, C.F., and LOHMAN, T.M. (1978). Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Quart. Rev. Biophys.* **11**: 103-178.

ROCA, A.I., and COX, M.M. (1990). The RecA protein: structure and function. *Crit. Rev. Biochem. Mol. Biol.* **25**: 415-456.

RUSCHE, J.R., KONIGSBERG, W., and HOWARD-FLANDERS, P. (1985). Isolation of altered recA polypeptides and interaction with ATP and DNA. *J. Biol. Chem.* **260**: 949-955.

RUSSEL, M., KIDD, S., and KELLEY, M.R. (1986). An improved filamentous helper phage for generating single-stranded plasmid DNA. *Gene* **45**: 333-338.

RUSSEL, M. (1991). Filamentous phage assembly. *Mol. Microbiol.* **5**: 1607-1613.

RUVOLO, P.P., KEATING, K.M., WILLIAMS, K.R., and CHASE, J.W. (1991). Single-stranded DNA binding proteins (SSBs) from prokaryotic transmissible plasmids. *Proteins: Struct. Funct. Genet.* **9**: 120-134.

SALAS, M. (1988). Initiation of DNA replication by primer proteins: bacteriophage ϕ29 and its relatives. *Curr. Top. Microbiol. Immun.* **136**: 71-88.

SALSTROM, J.S., and PRATT, D. (1971). Role of coliphage M13 gene 5 in single-stranded DNA production. *J. Mol. Biol.* **61**: 489-501.

SAMBROOK J., FRITSCH, E.F., and MANIATIS, T. (1989). *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

SANCAR, A., STACHELEK, C., KONIGSBERG, W., and RUPP, W.D. (1980). Sequences of the *recA* gene and protein. *Proc. Natl. Acad. Sci. USA* **77**: 2611-2615.

SANCAR, A., WILLIAMS, K.R., CHASE, J.W., and RUPP, W.D. (1981). Sequences of the *ssb* gene and protein. *Proc. Natl. Acad. Sci. USA* **78**: 4274-4278.

SANG, B.-C., and GRAY, D.M. (1989). Specificity of the binding of fd gene 5 protein to polydeoxyribonucleotides. *J. Biomol. Struct. Dynam.* **7**: 693-706.

SANGER, F., NICKLEN, S., and COULSON, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463-5467.

SCHALLER, H., UHLMANN, A., and GEIDER, K. (1976). A DNA fragment from the origin of single-strand to double-strand DNA replication of bacteriophage fd. *Proc. Natl. Acad. Sci. USA* **73**: 49-53.

SCHALLER, H., BECK, E., and TAKANAMI, M. (1978). Sequence and regulatory signals of the filamentous phage genome. In: *The Single-Stranded DNA Phages* (D.T. Denhardt, D. Dressler, and D.S. Ray, eds.) pp 139-163, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

SCHEIN, C.H. (1989). Production of soluble recombinant proteins in bacteria. *Bio/Technology* **7**: 1141-1149.

SCHMITT, M.P., BECK, P.J., KEARNEY, C.A., SPENCE, J.L., DIGIOVANNI, D., CONDREAY, J.P., and MOLINEUX, I.J. (1987). Sequence of a conditionally essential region of bacteriophage T3, including the primary origin of DNA replication. *J. Mol. Biol.* **193**: 479-495.

SCHWABE, J.W.R., and RHODES, D. (1991). Beyond zinc fingers: steroid hormone receptors have a novel structural motif for DNA recognition. *TIBS* **16**: 291-296.

SCHWARTZ, R.M., and DAYHOFF, M.O. (1978). Matrices for detecting distant relationships. In: *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 3 (M.O. Dayhoff, ed.) pp. 353-358, National Biomedical Research Foundation, Washington D.C.

SCHWARZ, G., and WATANABE, F. (1983). Thermodynamics and kinetics of co-operative proteins-nucleic acid binding. I. General aspects of analysis of data. *J. Mol. Biol.* **163**: 467-484.

SHAMOO, Y., GHOSAINI, L.R., KEATING, K.M., WILLIAMS, K.R., STURTEVANT, J.M., and KONIGSBERG, W.H. (1989). Site-specific mutagenesis of T4 gene 32: the role of tyrosine residues in protein-nucleic acid interactions. *Biochemistry* **28**: 7409-7417.

SHIEH, G.-J., CHARNG, Y.-C, YANG, B.-C., TU, J., BAU, H.-J., and KUO, T.-T. (1991). Identification and nucleotide sequence analysis of an open reading frame involved in high-frequency conversion of turbid to clear plaque mutants of filamentous phage Cf1t. *Virology* **185**: 316-322.

SHINE, J., and DALGARNO, L. (1974). The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71**: 1342-1346.

SHOEMAN, R.L., WADLE, S., SCHERBARTH, A., and TRAUB, P. (1988). The binding *in vitro* of the

intermediate filament protein vimentin to synthetic oligonucleotides containing telomere sequences. *J. Biol. Chem.* **263**: 18744-18749.

SILVER, P., and WICKNER, W.T. (1980). Role of M13 gene 1 protein in filamentous virus assembly. *J. Virol.* **35**: 256-258.

SIMONS, G.F.M., KONINGS, R.N.H., and SCHOENMAKERS, J.G.G. (1981). Genes VI, VII, and IX of phage M13 code for minor capsid proteins of the virion. *Proc. Natl. Acad. Sci. USA* **78**: 4194-4198.

SKINNER, M.M., ZHANG, H., LESCHNITZER, D.H., GUAN, Y., BELLAMY, H., SWEET, R.M., GRAY, C.W., KONINGS, R.N.H., WANG, A.H.J., and TERWILLIGER, T.C. (1994). Structure of the gene V protein of bacteriophage f1 determined by multiwavelength X-ray diffraction on the selenomethionyl protein. *Proc. Natl. Acad. Sci. USA*, in press.

SMITH, T.F., and WATERMAN, M.S. (1981). Comparison of biosequences. *Adv. Appl. Math.* **2**: 482-489.

SMITS, M.A., SIMONS, G., KONINGS, R.N.H., and SCHOENMAKERS, J.G.G. (1978). Expression of bacteriophage M13 DNA *in vivo*. I. Synthesis of phage-specific RNA and protein in minicells. *Biochim. Biophys. Acta* **521**: 27-44.

SMITS, M.A., SCHOENMAKERS, J.G.G., and KONINGS, R.N.H. (1980). Expression of bacteriophage M13 DNA *in vivo*. Isolation, identification and characterization of phage-specific mRNA species. *Eur. J. Biochem.* **112**: 309-321.

SMITS, M.A. (1982). Expression of bacteriophage M13 DNA. The *in vivo* transcription process. Ph.D. Thesis, University of Nijmegen, the Netherlands.

SMITS, M.A., JANSEN, J., KONINGS, R.N.H., and SCHOENMAKERS, J.G.G., (1984). Initiation and termination signals for transcription in bacteriophage M13. *Nucleic Acids Res.* **12**: 4071-4081.

STADEN, R. (1980). A new computer method for the storage and manipulation of DNA gel reading data. *Nucleic Acids Res.* **8**: 3673-3694.

STADEN, R. (1982). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Res.* **10**: 2951-2961.

STANISICH, V.A. (1974). The properties and host range of male-specific bacteriophages of *Pseudomonas aeruginosa*. *J. Gen. Microbiol.* **84**: 332-342.

STORY, R.M., WEBER, I.T., and STEITZ, T.A. (1992). The structure of the *E. coli* recA protein monomer and polymer. *Nature* **355**: 318-325.

SUGIMOTO, K., SUGISAKI, H., OKAMOTO, T., and TAKANAMI, M. (1977). Studies on bacteriophage fd DNA. IV. The sequence of messenger RNA for the major coat protein gene. *J. Mol. Biol.* **111**: 487-507.

TAKEYA, K., and AMAKO, K. (1966). A rod-shaped *Pseudomonas* phage. *Virology* **28**: 163-165.

TERWILLIGER, T.C., FULFORD, W.D., and ZABIN, H.B. (1988). A genetic selection for temperature-sensitive variants of the gene V protein of bacteriophage f1. *Nucleic Acids Res.* **16**: 9027-9039.

TIRANTI, V., BARAT-GUERIDE, M., BIJL, J., DIDONATO, S., and ZEVIANI, M. (1991). A full-length cDNA encoding a mitochondrial DNA-specific single-stranded DNA binding protein from *Xenopus laevis*. *Nucleic Acids Res.* **19**: 4291.

TORBET, J., GRAY, D.M., GRAY, C.W., MARVIN, D.A., and SIEGRIST, H. (1981). Structure of the fd DNA-gene 5 protein complex in solution. A neutron small-angle scattering study. *J. Mol. Biol.* **146**: 305-320.

TRAUB, P., MOTHES, E., SHOEMAN, R., KÜHN, S., and SCHERBARTH, A. (1992). Characterization of the nucleic acid-binding activities of the isolated amino-terminal head domain of the intermediate filament protein vimentin reveals its close relationship to the DNA-binding regions of some prokaryotic single-stranded DNA-binding proteins. *J. Mol. Biol.* **228**: 41-57.

TSUGITA, A., and KNEALE, G.G. (1985). Identification of lysine residues at the binding site of bacteriophage-Pf1 DNA-binding protein. *Biochem. J.* **228**: 193-199.

VOLODIN, A.A. (1990). Interaction with DNA and aggregation properties of the C-terminal fragment of the RecA protein. *Mol. Biol.* (Mosk) **24**: 179-188.

WATANABE, F., and SCHWARZ, G. (1983). Thermodynamics and kinetics of co-operative protein-nucleic acid binding. II. Studies on the binding between protamine and calf thymus DNA. *J. Mol. Biol.* **163**: 485-498.

WATANABE, T., NISHIDA, H., OGATA, C., ARAI, T., and SATO, S. (1964). Episome-mediated transfer of drug resistance in Enterobacteriaceae. VII. Two types of naturally occurring R factors. *J. Bacteriol.* **88**: 716-726.

WEBSTER, R.E., and REMENTER, M. (1980). Replication of bacteriophage f1: a complex containing gene II protein in gene V mutant-infected bacteria. *J. Mol. Biol.* **139**: 393-405.

WEBSTER, R.E., GRANT, R.A., and HAMILTON, L.A.W. (1981). Orientation of the DNA in the filamentous bacteriophage f1. *J. Mol. Biol.* **152**: 357-374.

WEBSTER, R.E., and LOPEZ, J. (1985). Structure and assembly of the class I filamentous bacteriophage. In: *Virus Structure and Assembly* (S. Casjens, ed.) pp 235-267, Jones and Bartlett Publishers, Boston.

VAN WEZENBEEK, P.M.G.F., HULSEBOS, T.J.M., and SCHOENMAKERS, J.G.G. (1980). Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd. *Gene* **11**: 129-148.

WICKNER, W., BRUTLAG, D., SCHEKMAN, R., and KORNBERG, A. (1972). RNA synthesis initiates *in vitro* conversion of M13 DNA to its replicative form. *Proc. Natl. Acad. Sci. USA* **69**: 965-969.

WICKNER, W., and KORNBERG, A. (1974). A holoenzyme form of deoxyribonucleic acid polymerase III. Isolation and properties. *J. Biol. Chem.* **249**: 6244-6249.

WICKNER, W., and KILLICK, T. (1977). Membrane-associated assembly of M13 phage in extracts of virus-infected *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **74**: 505-509.

WILBUR, W.J., and LIPMAN, D.J. (1983). Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA* **80**: 726-730.

WILLIAMS, K.R., LOPRESTI, M.B., and SETOGUCHI, M. (1981). Primary structure of the bacteriophage T4 DNA helix-destabilizing protein. *J. Biol. Chem.* **256**: 1754-1762.

WOOLFORD, J.L., STEINMAN, H.M., and WEBSTER, R.E. (1977). Adsorption protein of bacteriophage f1: solubilization in deoxycholate and localization in the f1 virion. *Biochemistry* **16**: 2694-2700.

WÜTHRICH, K. (1986). *NMR of Proteins and Nucleic Acids*, Chapter 1, and 7-10. Wiley and Sons, New York.

YANISCH-PERRON, C., VIEIRA, J., and MESSING, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103-119.

YEN, T.S.B., and WEBSTER, R.E. (1982). Translational control of bacteriophage f1 gene II and gene X proteins by gene V protein. *Cell* **29**: 337-345.

YOSHIKAWA, H., and ITO, J. (1982). Nucleotide sequence of the major early region of bacteriophage ϕ29. *Gene* **17**: 323-335.

ZABIN, H.B., and TERWILLIGER, T.C. (1991). Isolation and *in vitro* characterization of temperature-sensitive mutants of the bacteriophage f1 gene V protein. *J. Mol. Biol.* **219**: 257-275.

ZAMAN, G.J.R., SCHOENMAKERS, J.G.G., and KONINGS, R.N.H. (1990). Translational regulation of M13 gene II protein by its cognate single-stranded DNA binding protein. *Eur. J. Biochem.* **189**: 119-124.

ZAMAN, G.J.R., SMETSERS, A., KAAN, A.M., SCHOENMAKERS, J.G.G., and KONINGS, R.N.H. (1991). Regulation of expression of the genome of bacteriophage M13. Gene V protein regulated translation of the mRNAs encoded by genes I, III, V and X. *Biochim. Biophys. Acta* **1089**: 183-192.

ZAMAN, G.J.R. (1991). Translational regulation by the single-stranded DNA binding protein of the filamentous bacteriophage M13. Ph.D. Thesis, University of Nijmegen, the Netherlands.

ZEHNER, Z.E., LI, Y., ROE, B.A., PATERSON, B.M., and SAX, C.M. (1987). The chicken vimentin gene. Nucleotide sequence, regulatory elements, and comparison to the hamster gene. *J. Biol. Chem.* **262**: 8112-8120.

ZINDER, N.D., VALENTINE, R.C., ROGER, M., and STOECKENIUS, W. (1963). f1, a rod-shaped male-specific bacteriophage that contains DNA. Virology **20**, 638-640.

ZINDER, N.D., and HORIUCHI, K. (1985). Multiregulatory element of filamentous bacteriophages. *Microbiol. Rev.* **49**: 101-106.

SUMMARY

The interactions between proteins and nucleic acids are essential for the expression and replication of the genetic information. The proteins that interact with nucleic acids encompass both double-stranded DNA and single-stranded DNA binding proteins. Most detailed information is available for the double-stranded DNA binding proteins, while little is known about the proteins that bind to single-stranded DNA. This thesis deals with the single-stranded DNA binding protein encoded by gene V (gVp) of the filamentous *Escherichia coli* bacteriophage M13. A concise survey of what is known about the biophysical characteristics and biological functions of gVp is presented in Chapter 1. In the Appendix of this chapter a brief overview of the F-specific bacteriophages is given. After infection, the single-stranded viral genome is converted by host factors to double-stranded replicative form I. Transcription and subsequent translation of these molecules results in the synthesis of ten phage proteins. One of these, *i.e.* the site-specific topoisomerase encoded by gene II, initiates the synthesis of new viral strands by rolling-circle replication. Early after infection these strands are again converted to replicative form I resulting in a steady increase of the pool of these molecules as well as of the phage encoded proteins. Late in infection, when the gVp concentration has reached a critical threshold level, this protein binds co-operatively to the newly synthesized viral strand, thereby sequestering the viral genome from the replication machinery. In a subsequent stage the gVp molecules of the gVp•viral genome complex are exchanged for the coat proteins at the host cell membrane. Concomitantly with this exchange the virus particle is extruded from the cell. Besides its 'switch-function' in the process of viral genome replication, gVp also regulates the phage replication indirectly, *i.e.* via regulation of the expression of several phage genes at the level of translation by specific binding to the 5' non-translated leader of the mRNAs.

The goal of the studies described in this thesis was to investigate the structural characteristics and the relationship between structure and functions of this protein. In particular, which amino acids, or domains of gVp are involved in the execution of its distinct functions were studied.

To elucidate which amino acid residues or domains of gVp are involved in the various functions of the protein, the studies described in Chapter 2 were undertaken. Gene V was inserted in a phagemid expression vector and a library of mutants was constructed by random chemical mutagenesis. Phagemids encoding gene V proteins with decreased biological activities were selected and the nucleotide sequences of their gene V fragments were determined. Furthermore, the mutant proteins were characterized both with respect to their ability to inhibit the production of phagemid DNA transducing particles and their ability to repress the translation of a chimeric *lacZ* reporter gene whose expression is controlled by the promoter and translational initiation signals of M13 gene II. From the data obtained, it can be deduced that the mechanism by which gene V protein binds to single-stranded DNA differs from the mechanism by which it binds to its target sequence in the gene II mRNA.

The studies on the binding characteristics of mutant gVp containing single-site amino acid substitutions, are presented in Chapter 3. The series of mutant proteins tested includes mutations in the purported monomer-monomer interaction region as well as mutations in the DNA binding domain at positions which are thought to be functionally involved in monomer-monomer interaction or single-stranded DNA binding. The characteristics of the binding of the mutant proteins to the homopolynucleotides poly(dA), poly(dU), and poly(dT) were studied by means of fluorescence titration experiments. The binding stoichiometry and fluorescence quenching of the mutant proteins are equal to, or lower than, the wild-type gene V protein values. In addition, all the proteins measured bind in a more or less co-operative manner to single-stranded DNA. The binding affinities for poly(dA) decrease in the following order: Tyr61-His > wild-type > Phe68-Leu and Arg16-His > Tyr41-Phe and Tyr41-His > Phe73-Leu > Arg21-Cys > Tyr34-His > Gly18-Asp/Tyr56-His. The conservation of binding affinity, also for mutations in the single-stranded DNA binding domain, suggests that the binding to homopolynucleotides is largely non-specific and primarily dictated by electrostatic interactions.

The ^{1}H-NMR experiments on gVp to elucidate which aromatic amino acid residues are involved in single-stranded DNA binding, are described in Chapter 4. Using two-dimensional total correlation spectroscopy, sequence-specific assignments were obtained for all of the aromatic amino acid residues with the help of a series of single-site mutant proteins. Spin-labelled oligonucleotide-binding studies of wild-type and mutant gVp indicated that tyrosine 26 and phenylalanine 73 are the only aromatic residues involved in binding to short stretches of single-stranded DNA. The data obtained suggest the occurrence of specific protein-protein interactions between dimeric gVp molecules in which the tyrosine residue at position 41 is involved. This hypothesis is further strengthened by the observation that the solubility of tyrosine 41 mutants of gVp is significantly higher than that of the wild-type protein. The discovery of the so-called 'solubility' mutants of gVp has finally made it possible to elucidate the solution structure of gVp and its interaction with single-stranded DNA by means of two and three-dimensional ^{1}H-NMR.

Comparison of amino acid sequences of evolutionarily related proteins, which have comparable biological functions, might provide important information about the significance of particular amino acid residues or sequences with respect to the functions of the protein. Because the filamentous *E. coli* bacteriophage I2-2 has been suggested to be a naturally occurring host range mutant of phage IKe, comparison of the amino acid sequence of its ssDNA binding protein with those of phages Ff (M13, f1, fd) and IKe might be very informative. Therefore, the nucleotide sequence of the circular single-stranded genome of phage I2-2 was determined, and compared it with those of the filamentous *E. coli* phages IKe and Ff. The results of these studies are described in Chapter 5. The comparisons indicate that I2-2, IKe and Ff have a similar genetic organization, and that the genomes of I2-2 and IKe are evolutionarily more closely related than those of I2-2 and Ff. It is shown that the I2-2 genome is a composite replicon, composed of only two thirds of the ancestral genome of IKe. Only a contiguous I2-2 DNA sequence of 4615 nucleotides encompassing not only the coat protein and phage assembly genes, but also the signal required for efficient phage morhogenesis, was found to be significantly homologous to sequences in the genomes of IKe and Ff. No homology was observed between the consecutive DNA sequence that contains the origins for viral and complementary-strand replication and the replication genes. Although other explanations cannot be ruled out, our data strongly suggest that the ancestor filamentous phage genome of phages I2-2 and IKe has exchanged its replication module during evolution with that of another replicon, *e.g.* a plasmid that also replicates via the so-called rolling-circle mechanism.

^{1}H-NMR studies of single-stranded DNA binding proteins of the evolutionarily related filamentous *E. coli* phages Ff and IKe have shown that these proteins contain a structural motif, the so-called DNA binding wing, that intimately is involved in their non-sequence-specific binding to single-stranded DNA. To obtain a first indication whether this DNA binding wing is a recurrent motif in other single-stranded DNA binding proteins, an analysis was performed to investigate whether the amino acid sequences of these wings are recurrent sequences in other known prokaryotic and eukaryotic single-stranded DNA binding proteins. The analyses revealed that of all ssDNA binding proteins whose amino acid sequence has been stored in the protein databases, only ssDNA binding proteins encoded by the bacteriophages Ff, IKe, Pf1, Pf3, ϕ29, If1, I2-2, and T4 plus the eukaryotic intermediate filament protein vimentin, comprise an amino acid sequence that exhibits a significant degree of homology with this sequence.

SAMENVATTING

De interactie tussen eiwitten en nucleïnezuren is essentieel voor de expressie en vermenigvuldiging van erfelijke informatie. De eiwitten die een interactie met nucleïnezuren aangaan, omvatten zowel dubbel- als enkelstrengs DNA bindende eiwitten. Terwijl over de dubbelstrengs DNA bindende eiwitten veel gedetailleerde informatie beschikbaar is, is van de enkelstrengs DNA bindende eiwitten maar weinig bekend. Dit proefschrift handelt over het enkelstrengs DNA bindende eiwit dat gecodeerd wordt door het gen V (het zogenaamde gen V eiwit) van de draadvormige bacteriofaag M13 die de darmbacterie *Escherichia coli* als gastheer heeft. In Hoofdstuk 1 wordt een beknopt overzicht gegeven van de biofysische eigenschappen en de biologische functies van het eiwit. In de Appendix van dit hoofdstuk wordt in het kort een overzicht gegeven van de F-specifieke bacteriofagen waartoe M13 behoort. Na infectie wordt het enkelstrengs virale genoom, met behulp van eiwitten van de gastheercel, omgezet in een dubbelstrengs replicatieve vorm. Transcriptie en de daaropvolgende vertaling van deze moleculen resulteert in de synthese van tien faageiwitten. Eén daarvan, het plaatsgerichte topoisomerase dat door gen II gecodeerd wordt, initieert de aanmaak van nieuwe virale strengen door middel van het "rollende cirkel" DNA vermenigvuldigingsmechanisme. Aanvankelijk worden de nieuw gevormde virale strengen omgezet in de dubbel-strengs replicatieve vorm, hetgeen een ophoping van dit DNA alsmede een ophoping van faageiwitten tot gevolg heeft. Later, als de concentratie gen V eiwit een bepaalde drempelwaarde heeft bereikt, bindt het gen V eiwit aan de nieuwe enkelstrengs DNA moleculen. Deze binding gebeurt op een coöperatieve manier, dat wil zeggen dat binding van één gen V eiwit molecuul de binding van het volgende direct ernaast vergemakkelijkt. Ten gevolge van de binding van het gen V eiwit aan de nieuwe enkelstrengs DNA moleculen, worden deze niet meer omgezet in de dubbelstrengs replicatieve vorm. Vervolgens gaat het complex van gen V eiwit en viraal enkelstrengs DNA naar de celmembraan, waar de gen V eiwitmoleculen uitgewisseld worden voor manteleiwitten, terwijl de nieuwe faag wordt uitgescheiden. Naast zijn functie als "schakelaar" in het DNA vermenigvuldigingsproces, reguleert het gen V eiwit de expressie van een aantal faaggenen op het niveau van de vertaling van de boodschapper RNAs.

Het doel van het onderzoek dat in dit proefschrift is beschreven, was de bestudering van de structurele eigenschappen èn de relatie tussen structuur en functie van het gen V eiwit. Daarbij stond de vraag welke aminozuren of domeinen in het gen V eiwit betrokken zijn bij de diverse functies, centraal.

Om te achterhalen welke aminozuren of gebieden van het gen V eiwit bij de verschillende functies van het eiwit betrokken zijn, werden de experimenten die in Hoofdstuk 2 beschreven staan, uitgevoerd. Gen V werd in een faagmiede expressie vector geïnserteerd en met behulp van chemische mutagenese werd een verzameling van mutanten geconstrueerd. Faagmieden die een eiwit codeerden met een verlaagde biologische activiteit werden geselecteerd, en de nucleotiden-volgorde van het gen V dat zij bevatten werd bepaald. Verder werd van de gemuteerde eiwitten gemeten in hoeverre ze in staat waren om de produktie van faagmiede DNA bevattende partikels te remmen, en in hoeverre ze de vertaling van het fusie RNA, bestaande uit een gedeelte van M13 gen II en het β-galactosidase gen van *E. coli*, konden remmen. Uit de verkregen resultaten kon worden afgeleid dat het mechanisme van binding aan enkelstrengs DNA anders is dan het mechanisme van binding aan gen II boodschapper RNA.

In Hoofdstuk 3 worden de experimenten beschreven die gericht waren op de bepaling van de bindingseigenschappen van mutante gen V eiwitten met één aminozuursubstitutie. Deze mutante eiwitten bevatten zowel mutaties in het veronderstelde monomeer-monomeer interactie gebied als in het DNA bindend gebied, en wel op posities waarvan vermoed wordt dat ze betrokken zijn bij monomeer-monomeer interacties of binding aan enkelstrengs DNA. Met behulp van fluorescentietitratie-experimenten zijn de bindingskarakteristieken van de binding van mutante eiwitten aan de homopolynucleotiden poly(dA), poly(dU) en poly(dT) bestudeerd. De stoichiometrie van de binding en de uitdoving van de fluorescentie van de mutante eiwitten zijn gelijk of lager dan die van het wildtype eiwit. Bovendien binden alle eiwitten in meer of mindere mate coöperatief aan enkelstrengs DNA. De affiniteit voor poly(dA) neemt af in de volgorde: Tyr61-His > wildtype > Phe68-Leu en Arg16-His > Tyr41-Phe en Tyr41-His > Phe73-Leu > Arg21-Cys > Tyr34-His > Gly18-Asp/Tyr56-His. De experimenten laten zien dat de affiniteit voor binding behouden blijft, ook voor mutanten met veranderingen in het DNA bindend gebied. Dit duidt er op dat de binding aan homopolynucleotiden voornamelijk niet-specifiek is en met name door electrostatische interacties bepaald wordt.

De ^{1}H-NMR experimenten die tot doel hadden te onderzoeken welke aromatisch aminozuren bij de binding aan enkelstrengs DNA betrokken zijn, staan beschreven in Hoofdstuk 4. Met behulp van tweedimensionale ^{1}H-NMR technieken werden sequentie specifieke toekenningen verkregen door gebruik te maken van enkele mutante gen V eiwitten waarin één aromatisch aminozuur was vervangen door een ander aminozuur. Experimenten met oligonucleotiden die een zogenaamd spin-label bevatten, hebben aangetoond dat van de aromatisch aminozuren in het gen V eiwit, alleen tyrosine 26 en fenylalanine 73 bij de

binding aan enkelstrengs DNA betrokken zijn. De resultaten duiden er verder op, dat er specifieke eiwit-eiwit interacties tussen dimeren van gen V eiwit optreden waarbij tyrosine 41 betrokken is. Deze vooronderstelling wordt onderbouwd door de waarneming dat de oplosbaarheid van mutante eiwitten, waarin tyrosine 41 vervangen is door een ander aminozuur, significant hoger is dan van het wild-type eiwit. De ontdekking van deze zogenaamde "oplosbaarheid" mutanten heeft het uiteindelijk mogelijk gemaakt om de structuur van het gen V eiwit in oplossing en de interactie van dit eiwit met enkelstrengs DNA met behulp van twee- en drie-dimensionale ^{1}H-NMR te bestuderen.

De vergelijking van aminozuursequenties van evolutionair verwante eiwitten met vergelijkbare biologische functies, kan belangrijke informatie opleveren wat betreft de betekenis van bepaalde aminozuren of aminozuursequenties. Omdat de draadvormige bacteriofaag I2-2 verwant is aan de draadvormige bacteriofaag IKe, kan de vergelijking van de enkelstrengs DNA bindende eiwitten van de fagen M13, IKe en I2-2 informatief zijn. Daarom hebben we de totale nucleotiden-volgorde van het genoom van de faag I2-2 bepaald en vervolgens vergeleken met die van de fagen IKe en Ff. De resultaten van deze experimenten zijn beschreven in Hoofdstuk 5. De vergelijking van de nucleotiden-volgorden wees uit dat I2-2, IKe en Ff dezelfde genetische organisatie hebben, en dat de genomen van I2-2 en IKe evolutionair nauwer verwant zijn dan die van I2-2 en Ff. Verder is aangetoond dat het I2-2 genoom een composiet replicon is, dat slechts weederde van het voorouderlijke genoom van IKe bevat. Een aaneengesloten I2-2 nucleotiden-volgorde van 4615 nucleotiden, die de manteleiwit- en assemblage genen, alsmede signalen die nodig zijn voor correcte morfogenese, bevat, is significant homoloog met nucleotiden-volgorden van de fagen IKe en Ff. Daarentegen is er geen homologie tussen de nucleotiden-volgorden die de vermenigvuldigingsgenen en de beginpunten van de DNA vermenigvuldiging bevatten. Ofschoon andere verklaringen niet uitgesloten zijn, wijzen onze gegevens er sterk op dat het voorouderlijke draadvormige faaggenoom van de fagen I2-2 en IKe gedurende de evolutie een vermenigvuldigingsmodule met een ander replicon, bijvoorbeeld een plasmiede dat ook volgens het "rollende cirkel" DNA vermenigvuldigingsmechanisme repliceert, heeft uitgewisseld.

^{1}H-NMR experimenten met de enkelstrengs DNA bindende eiwitten van de evolutionair verwante *E. coli* fagen IKe en Ff, hebben aangetoond dat deze eiwitten een structureel motief bevatten dat betrokken is bij de niet specifieke binding van deze eiwitten aan enkelstrengs DNA. Om een indruk te verkrijgen of dit motief een terugkerend fenomeen bij enkelstrengs DNA bindende eiwitten is, werd de computer-analyse uitgevoerd zoals in Hoofdstuk 6 beschreven. Van alle bestudeerde prokaryotische en eukaryotische enkelstrengs DNA bindende eiwitten, bevatten alleen de enkelstrengs DNA bindende eiwitten van de bacteriofagen Ff, IKe, Pf1, Pf3, ϕ29, If1, I2-2 en T4, alsmede het eukaryotische cytoskeleteiwit vimentine, een aminozuursequentie die significant homoloog is met die van het gen V eiwit van de fagen IKe en M13.

LIST OF PUBLICATIONS

KELTJENS, J.T., VAN BEELEN, P., STASSEN, A.P.M., and VOGELS, G.D. (1983). 7-Methylpterin in methanogenic bacteria. *FEMS Microbiol. Lett.* **20**: 259-262.

VAN BEELEN, P., STASSEN, A.P.M., BOSCH, J.W.G., VOGELS, G.D., GUIJT, W., and HAASNOOT, C.A.G. (1984). Elucidation of the structure of methanopterin, a coenzyme from *Methanobacterium thermoautotrophicum*, using two-dimensional nuclear-magnetic-resonance techniques. *Eur. J. Biochem.* **138**: 563-571.

SMITS, R.A.M.M., VAN DE WIJNGAARD, W.M.H., STASSEN, A.P.M., and VAN DER DRIFT, C. (1984). Mutants of *Pseudomonas aeruginosa* unable to inactivate allantoinase and NADP-dependent glutamate dehydrogenase. *Arch. Microbiol.* **140**: 40-43.

VAN DUYNHOVEN, J.P.M., FOLKERS, P.J.M., STASSEN, A.P.M., HARMSEN, B.J.M., KONINGS, R.N.H., and HILBERS, C.W. (1990). Structure of the DNA binding wing of the gene-V encoded single-stranded DNA binding protein of the filamentous bacteriophage M13. *FEBS Lett.* **261**: 1-4.

FOLKERS, P.J.M., STASSEN, A.P.M., VAN DUYNHOVEN, J.P.M., HARMSEN, B.J.M., KONINGS, R.N.H., and HILBERS, C.W. (1991). Characterization of wild-type and mutant M13 gene V proteins by means of ^{1}H-NMR. *Eur. J. Biochem.* **200**: 139-148.

STASSEN, A.P.M., ZAMAN, G.J.R., VAN DEURSEN, J.M.A., SCHOENMAKERS, J.G.G., and KONINGS, R.N.H. (1992). Selection and characterization of randomly produced mutants of gene V protein of bacteriophage M13. *Eur. J. Biochem.* **204**: 1003-1014.

STASSEN, A.P.M., SCHOENMAKERS, E.F.P.N., YU, M., SCHOENMAKERS, J.G.G., and KONINGS, R.N.H. (1992). Nucleotide sequence of the genome of the filamentous bacteriophage I2-2: module evolution of the filamentous phage genome. *J. Mol. Evol.* **34**: 141-152.

STASSEN, A.P.M., HARMSEN, B.J.M., SCHOENMAKERS, J.G.G., HILBERS, C.W., and KONINGS, R.N.H. (1992). Fluorescence studies of the binding of bacteriophage M13 gene V mutant proteins to polynucleotides. *Eur. J. Biochem.* **206**: 605-612.

CURRICULUM VITAE

De schrijver van dit proefschrift werd geboren op 29 oktober 1958 te Heerlen. Na het behalen van het MAVO- (1975), HAVO- (1977) en VWO- (1978) diploma, begon hij in september 1978 met de studie Biologie aan de Katholieke Universiteit Nijmegen. In september 1981 werd het kandidaatsexamen B4 met succes behaald. Het doctoraal-examen werd in november 1981 *met genoegen* afgelegd. Dit examen omvatte het bijvak Chemische Cytologie (Prof. Dr. C.M.A. Kuyper) en de hoofdvakken Microbiologie (Prof. Dr. Ir. G.D. Vogels) en Moleculaire Biologie (Prof. Dr. J.G.G. Schoenmakers). Vanaf 1 januari 1986 tot en met 31 december 1989 was hij als wetenschappelijk assistent verbonden aan het Laboratorium voor Moleculaire Biologie van de Katholieke Universiteit Nijmegen. Tijdens deze periode werd onder leiding van Prof. Dr. R.N.H. Konings het in dit proefschrift beschreven onderzoek verricht.

Momenteel werkt hij als wetenschappelijk medewerker bij de Sectie Moleculaire Genetica van het Nederlands Kanker Instituut te Amsterdam.